

Rethinking readability of digital editions – the case of the AAC’s “Digital Brenner”

Karlheinz Mörth

Austrian Academy Corpus
Sonnenfelsgasse 19, 1010 Vienna, Austria
karlheinz.moerth@oeaw.ac.at

Abstract

Retrodigitization projects have become very common as demand for digital versions has increased considerably over the past few years. Many of these projects have been limited to the production of facsimiles, to a much lesser degree text has been integrated which is due to the costs that are involved in such undertakings. The AAC’s “Brenner” project was started as an experimental endeavour in humanities computing designed to fathom out the possibilities involved in the particular case of historical journals. The focus in this paper is on issues of readability of digital texts. The issue at hand is the exploration of features of digitized editions that may make up for the printed original and even go beyond what is offered by the printed original. The identification of useful display objects that can be easily implemented digitally was the main concern of the study. In describing the constituents of a planned web interface, we will try to figure out ways of improving the acceptability and quality of electronic texts.

Prologue

It is a truism that images have come to play an ever more important role in both electronic and print media. However, reading is still one of the most essential cultural techniques in information society. Perceiving and comprehending textual data is the main gate to any kind of higher level knowledge. People read for various reasons and in different ways. Although computers have been available for some time now, the number of people who read large amounts of text from monitors is still low. Attempts to replace printed texts by digital texts have so far failed in many areas. After all, reading on the screen is still impeded by factors related to the highly complex nature of the reading process as such.

Along with widespread eye-fatigue, the lack of physical mobility of the reader is beyond doubt one of the most obvious obstacles to extensive and yet comfortable reading of electronic texts. One can read a book almost anywhere or in any position: not so with the computer. Practice has shown that even professional computer workers often print texts out in order to work on them. Actually, text production by electronic means turned out to yield far more paper than it did in times when there were no computers around. Even laptops have not made up for this physical restriction and the most recent generation of tablet PCs does not seem to gain ground as quickly as expected.

In general, electronic texts are considered less readable than their printed counterparts. Apart from hardware deficiencies, a number of other problems can be discerned. Textual electronic products often do not display many of the evident possibilities of the new medium. Poorly structured and inadequately adapted electronic replica of previously printed material are nowadays ubiquitous.

So-called electronic texts fall into two major categories: digitized texts, which are electronic replications of previously printed material, and texts that have no printed analogue. The degree of readability in the second category is generally better, as these are already tailored to the demands of the new medium. What we are interested in here is the first category of texts, retrodigitized data, data

transferred from printed to digital resources. As the overwhelming majority of our literary heritage (literary is meant here in an extensive sense encompassing all sorts of written materials) was produced in pre-digital times, most of what people may want to read, or what scholars and scientists may need for their studies does not exist in digital form and is still awaiting transfer to the new medium.

Most texts dating from pre-digital times have not been digitised so far and maybe never will be. However, digital versions of printed materials are being produced in unprecedented numbers these days. The process of creating digital literary resources from printed originals always implies electronic editing, as the process necessarily requires a varying amount of textual reconstruction in the new medium. An often underestimated factor is the amount of both time and money involved in such undertakings. Digitising printed material is for the most part an immensely work-intensive and time-consuming task. The creation of digitized editions that can be used as the basis of scholarly work requires specialized knowledge and scholars willing and able to embark on such challenging projects.

Lack of readability has also been caused by the focus on retrieval, which may in turn be a result of the predominance of linguistics over other disciplines in the field of humanities computing. Online searching is the first thing that comes to mind, rather than extensive reading on screen. Corpora have come to be an accepted and widely used means of study in the domain of linguistic research, yet much less so in literary studies, history, sociology and other fields of the Humanities. Although literary texts are being made available in ever-increasing numbers these days, one has to be aware of the fact that many of these came into existence as by-products of linguistic efforts to accumulate data for lexicographic or learners’ purposes rather than as results of projects designed to expand the digital literary canon. A great number of the literary sources at hand do not meet the standards that are generally regarded as a prerequisite for scholarly research in this field.

Framework of the study

When you start looking for digital historical texts in the German language you will come across a number of sites focussing on history¹ and jurisprudence². Numerous important documents in these fields have been digitized in recent years. The situation does not look so good for researchers interested in literary data. Although material is available, we are by far not yet in a situation that allows us to access digitally even the most important German language literary works, let alone in acceptable editions. The foundation of so-called competence centers for humanities computing did not really increase digital output in our field of interest.

The Austrian Academy Corpus (AAC) as one such centre is a research unit set up by the Austrian Academy of Sciences, Austria's largest non-university research facility, which supports numerous scientific and scholarly projects across all fields of study on both a national and international level. The unit was founded with the aim of promoting and conducting theoretical and applied research in the field of electronic text corpora.

The Corpus

The AAC's main concern is digital literary texts. The Corpus has been designed as a modular system made up of a number of complementary subcorpora that can be structured and restructured according to varying demands. Besides thematic considerations, the type of text 'carriers' also plays an important role in the make-up of the corpus. Among the principles of the AAC's digitizing endeavours has been the preservation of textual integrity, striving for the digitization of coherent textual resources.

The startup phase of this scholarly corpus project has been characterized by work on historical literary and political journals. Among other projects one of the AAC's working groups is currently dealing with the digitization of historical newspapers, which also proved to be a rather challenging task. Yet another working group has started to build a small subcorpus that is designed to hold bilingual versions of prominent literary works together with relevant translations.

Technical considerations

The fascination with new media and its implications in the field of literary studies has been one of the team's main motives. How to encode texts, which markup languages to adopt, how to best deliver the results, these are the main issues the AAC has been concerned with in recent years. In fact, the AAC has been active in this field since the early nineties of the past century when PCs started to become popular.

A lot of effort has been spent on considering markup systems. Traditionally, there exists a dichotomy between systems that catered to the print industries, and those that grew out of archiving traditions. The interest in formatting information is obviously related, though not restricted, to the first group. There is little standardization in this sphere. The number of formats is great with many of them

being proprietary in nature and bound to specific software applications that are in turn tied to specific companies. Although the demarcation lines between operating systems seem to blur in recent years, many of these formats can neither be regarded as platform independent nor can one be sure of the accessibility of software to process this data.

Scholars, and not only in the arts and humanities, have generally worked in the archiving tradition where encoding has been dominated by two standards: the Standard Generalized Markup Language (SGML) and the Text Encoding Initiative (TEI). SGML is a markup language that was primarily designed to describe the structure of electronic documents. The TEI is a widely recognized group of people and supporting institutions, who have pushed for the establishment of consistent standards in the field of text encoding. The youngest descendant of the SGML derived language family is XML (Extensible Markup Language), basically a subset of SGML. Being stricter than SGML in terms of structure, XML and its sub-specifications allow scholars to give considerably more attention to the issue of format and also to other problems such as linking, querying etc., which so far have only been partly solved.

The AAC has committed itself to the application of XML in all its projects. It started to make use of this nowadays well-established technology in 1998, when barely any software was available which made necessary considerable in-house development work.

The "Digital Brenner" Project

Amongst the more than 50 literary and political journals in the AAC there is "Der Brenner", which has been described as an important document of recent Austrian cultural history. The journal was edited by the writer and essayist Ludwig von Ficker (1880-1967). The first issue appeared in June 1910. In the beginning the journal had a certain leaning towards expressionism. The first contributors were primarily authors from the Tyrol (Carl Dallago, Max von Esterle, Hugo Neugebauer, Karl Röck, Ludwig Seifert, Arthur von Wallpach), but very soon writers from other parts of the German-speaking area joined in (Theodor Haecker, Karl Borromäus Heinrich, Else Lasker-Schüler, Ludwig Erik Tesar). The poet Georg Trakl counts among the most prominent figures of the early period of the "Brenner". Having started as a platform of the literary avantgarde, the focus of the magazine continually shifted towards a more religious orientation during and after the years of World War I. The prominent names of this period are Ferdinand Ebner, Theodor Haecker (who translated Kierkegaard) and John Henry Cardinal Newman. After 1945 the "Brenner" appeared just three more times in separate volumes in the years 1946, 1948 and 1954.

A tentative typology – two categories of digital objects

Talking about the representation of digital textual data and what kinds of secondary data might be applied to enhance their readability, we will make use of the term *digital objects* in the very general sense of any kind of graphical entities serving as transactional anchors between the user and the underlying data structures, ranging from simple navigational buttons to text passages furnished with

¹ The World War I Document Archive (<http://www.lib.byu.edu/~rdh/wwi/>), EuroDocs: Primary Historical Documents From Western Europe.

² like [bundesgesetze.de](http://www.bundesgesetze.de) or the OSCE's Documents library (<http://www.osce.org/docs/>).

hyperlinks. The concern will be in identifying useful display objects rather than considering specific implementations.

It is not only the previously mentioned issues that can be blamed for the unwieldiness of many digital texts. Digital readability is an issue that goes far beyond the proper choice of fonts, font size and background colours. A number of rather more complex issues are at hand which need closer consideration. Creating digital editions is a by far more demanding task than is generally thought because in addition to the well-known problems of traditional print-editions, the scholar is confronted with a host of additional issues to be worked through.

What often gives readers an uncanny feeling when reading on screen has a number of reasons. One of the main drawbacks of many digital texts is the lack of evident 'anchoring' in space and time, aspects that relate to issues of navigation and user interface design in general.

When dealing with printed materials readers gain considerable information about what they have in 'hand' through physical contact with the artefact itself. The design of a book, of a magazine, of any printed item conveys information that allows a quick first judgment. The size of a physical reading object, a book or a journal, gives approximate data concerning the amount of text within, the design may give you clues as to genre, likely contents, may even allow speculations concerning the time or place of the objects publication.

To discuss the intricate interrelations of space and time is not the issue at hand, however the dependencies are obvious. Time is involved in several respects: picking up a book, a potential reader might be able to infer whether the specific book is old or new, very often they will be able to tell approximately when it was produced, how much time will be needed to read it and, having started to read it, how much time will be necessary to accomplish the task of reading.

In creating digital editions, scholars should keep thinking in two directions. The first perspective should be compensation for the above described deficiencies. What should be the essential constituents of digital editions that can make good the missing tactile and visual input of printed material? In addition to compensatory features, we have to think about how to achieve added value, how to furnish digital editions with information that goes beyond what is accessible in the printed version.

Compensatory features

In representing previously printed or handwritten materials on electronic devices we can discern two prototypical approaches: either the printed original is converted to digital text or the underlying text is reproduced in the form of digital facsimiles.

The first approach of downsizing the text to the character-level is very often pursued for rather pragmatic reasons³. Linguistic corpora do not usually need secondary data describing the appearance of the printed original. Therefore they reduce the digitised versions of such materials to 'pure' language data, leaving aside the whole

issue of layout and typography, concentrating rather on lemmas and parts of speech.

The situation is different, however, when it comes to editions of literary texts. The need for additional formal information is much more perceptible in such digitisation projects. This is why editors in this field often make use of images in order to preserve information concerning layout and typography which may not be regarded as mere decoration or ornament in such contexts. This approach is advisable especially when dealing with manuscripts or early prints⁴. When working along these lines, scholars have pursued different methods of relating the text to the image. Probably the most widespread approach is the juxtaposition of image and text in separated windows or frames⁵. An additional option is the linking of sections of the image to parts of the texts or annotations. This can be achieved through mechanisms such as image maps⁶.

Many electronic editions of literary works are implemented using a combination of facsimile and text. Unfortunately, interfaces are often not very user-friendly. The texts are accessible to professional researchers but admittedly do not really improve the degree of readability of the texts. It is difficult to navigate using facsimiles. Reading and finding correlating secondary textual information is often hampered by ineffective zooming and positioning mechanisms.

In addition, facsimiles do not offer direct access to the contents of the text itself introducing an additional layer which is only motivated by the electronic medium and as such superfluous. Immediate access to the underlying language data is often instrumental in further study. One might also consider highlighting parts of the texts (e.g. names, specific categories of parts of speech etc.) which further support scholarly analysis as well as having value in teaching and learning environments. Such features cannot easily be implemented when working only with facsimiles.

A third option when working with texts in the electronic medium falls somewhere in between: the emulation of the appearance of an original text combining formatting information with the textual data⁷. So far scholars have not made much use of this possibility.

We believe that in projects like the one we are discussing, only a dual approach can work. Images without the text

⁴ The *princeps* edition of *Don Quixote* (Madrid 1605) has been on the internet in form of images for some time (www.csdl.tamu.edu/cervantes/english/ctxt/princeps).

⁵ Examples are legion. A good example of such an approach is the electronic edition of Viktor Rydberg's Swedish translation of Johann Wolfgang von Goethe's *Faust* prepared as part of *Projekt Runeberg*. The transcription of the original is accompanied by images of the 5th edition of the translation (Stockholm 1915). For further details see www.lysator.liu.se/runeberg/faust.

A second example of such an adaptation is Marcel De Smedt and Edward Vanhoutte's very complex and highly sophisticated electronic-critical edition of Stijn Streuvel's *De teleurgang van den Waterhoek* (*The decline of the Waterhoek*), published on CD-Rom by Amsterdam University Press.

⁶ A well-known specimen of this kind is the *Electronic Beowulf* (2 CD-Roms published by British Library Publications and the University of Michigan Press).

⁷ A popular example on the internet is Masaru Uchida's electronic version of Laurence Sterne's *Tristram Shandy*, who attempted to emulate the original by means of HTML (<http://www.gifu-u.ac.jp/~masaru/TS/contents.html>).

³ An example of such an edition is Fred F. Jehle's adaptation of Miguel de Cervantes Saavedra's *Don Quixote de la Mancha* (www.ipfw.edu/cm1/jehle/web/cervante/dqsb1.htm).

will not work, nor the other way around. The implementation of an adequate graphical user interface did not pose too many problems in the case of the “Brenner” project as the textual composition of the printed original is rather straightforward. Real problems may arise with data of a more intricate structure or in large print format, which are more difficult to represent on electronic devices with their limited ‘surfaces’.

To enable readers to get their bearings within a digital text, it is necessary to furnish them with information as to the amount of data comprised in a specific digital object. When working with digital texts users more often than not do not know the amount of data involved. Even if a user is given the size in KBs or MBs, they will hardly be in a position to work out what that means in terms of real-world reading efforts. However, you would hardly embark on the task of reading a book without having an idea of how much time will be involved. In preparing digital texts we have to think more about ways to visualize this kind of information. One might conceive of different approaches, the most abstract of which is the numeric display of the number of pages and/or number of words, characters and such like values. We believe that the display of numbers will not be sufficient. They may even be quite misleading as the size of a digital text very much depends on the system of annotation and the depth of markup. Again, images could be a first step in the right direction, although images alone will not be enough. The problem with any display objects on screen is that they do not truly relate to the real world of the user. To be more than mere ornaments such digital objects will have to be provided with additional functionalities that allow comparisons of specific digital objects with objects known to the user. Scalar graphical representations of the data might be more user-friendly. Schematized boxes that symbolize this data or even scalar images of the books, journals or manuscripts are probably a more realistic approach to the issue.

To facilitate orientation within digital objects, special ‘landmarks’ symbolizing at which particular position a specific page is positioned within the digital object would be helpful. This kind of display object could furnish the reader with the information that the page they have been directed to by a search engine is at the beginning of the digital version of a text, in the middle or at the end. This feature of an interface should also be classified as compensatory as it attempts to emulate qualities always present in physical artefacts.

To fulfil a modicum of compensation for features found in printed data, a digital edition has also to provide linear navigation in form of sequential reciprocal links. This means that there have to be navigational mechanisms that allow forward and backward navigation. In addition, random access to any part of the digital object should be possible in an as comfortable manner as you would find in opening a book on any page you desire.

Added value

Symbols that help readers to figure out their current reading position within a specific digital object are evidently to be regarded as mere compensatory features. They simply make up for something that is available to everyone in possession of the physical object. Quite different is the case of display objects that relate the

current positioning to external parameters. Obviously these constitute a different category of navigational aids. One could conceive of timelines with indicators of the current position. Being told at any position of a text or a text collection, when the specific object was published might be a hypertext feature that clearly furnishes added value, value not inherent in the original print version. Such navigational objects may be used to visualize scalar values such as years or months. In other contexts they could establish relationships to historic events or persons. Navigational controls that allow moving from one page to the next and thus enable sequential reading of the text are also compensatory in nature. Most digital editions of literary works do not offer anything but limited hyperlinking. In the “Brenner” project we have tried to allow the user to move directly to specific points of the text, such as the beginning of the current article, the start of the next, the commencement of an issue, related contents pages, features that obviously go beyond mere compensation. Looking at all the beginnings of all the chapters in a book obviously involves sifting through the whole volume. Doing this in a hyperlinked edition is a much more feasible undertaking. Searching for the beginning of an article in a printed edition means looking at each and every page, a thoroughly annotated digital edition could provide this kind of point-to-point navigation as a standard feature. Further options that might easily be implemented are links to specific cognate positions, e.g. from one title to the following one, from one caricature to the next, from one contents page to the next.

Evidently searching capabilities count among those features that first come to mind when discussing how to furnish digitized versions with added value. However, the edit control that controls the search engine is only one possible access point to a digital edition.

In the web interface of the “Brenner” project we have implemented two additional features: (a) a list of all articles mentioned on the contents pages and (b) complete indices of all contributors (authors, translators, caricaturists). The listings are furnished with hyperlinks. All of these digital objects may be grouped in part under added value.

References

- Biber, Hanno & Breiteneder, Evelyn (2002). Austrian Academy Corpus: digital resources in textual studies. In *Digital Resources for the Humanities 2001–2002*. An edited selection of papers. Ed.: Jean Anderson, Alastair Dunning, Michael Fraser. Office for Humanities Communication 14 (pp. 13--18).
- Methlagl, Walter & Saueremann, Eberhard & Scheichl, Sigurd Paul (Ed., 1981). *Untersuchungen zum “Brenner”*. Festschrift für Ignaz Zangerle zum 75. Geburtstag. Salzburg.
- Mörth, Karlheinz (2002). The representation of literary texts by means of XML: some experiences of doing markup in historical magazines. In *Digital Evidence*. Selected papers from DRH 2000, Digital Resources for the Humanities Conference. Ed.: Michael Fraser, Nigel Williamson and Marilyn Deegan. Office for Humanities Communication 14 (pp. 17--32).
- Stieg, Gerald (1976). *Der Brenner und die Fackel*. Salzburg.