

ON-LINE CRITICAL EDITIONS IN THE 90S: PRACTICE AND THEORY

IVÁN HORVÁTH
ivan.horvath@elte.hu

Abstract - Every text is plural. A text is what can be copied exactly. Text digitizing is nothing else than a kind of exact copying. The digitization of images and sounds is nothing but their transformation into text. The more the accessibility of texts increases, the more text carrier durability is reduced, because accessibility is ensured by copying. A text only has to last until it is copied. The lesser amount of energy copying needs, the more probably it will occur. The advantage of text corruption: the corrupted text's fitness value changes. Environmental adaptation is what textual critics call lectio faciliior: the disappearance of obscure passages. After many a corruption, a new text comes into being. Textual criticism hinders adaptive modification.

Keywords - theory of literature, information technology

1. Digital technologies and philological disciplines

We are currently witnessing the third stage of information technology. At the beginning, there was the machine. One needed a soldering iron and a screwdriver to program it but at the time it was unable to perform more than one task. As long as the radar provided data concerning the direction, speed and height of the attacking airplane, this machine would calculate the right angle and the direction of the gun-barrel of the anti-aircraft canon, and would set the time of firing.

And then, there was the program. The computer architecture introduced by John von Neumann needed not only data input before it was able to operate, but also the set of instructions it had to carry out, since this was already a multi-purpose, universal device. With the intelligent programs running on the computers of our days, millions of computers, and even more users can be connected, still

they manage to retain a simple, user-friendly interface that anybody can easily learn to operate.

Finally, we have arrived at the age of content. We can now surf the sea of information, and browse the net with our wonderful computers that provide us with a user-friendly interface - only to realize that there is nothing there. Or, to be more precise, there is everything. Finally, we manage to fish out a text with our net browser, find a blue word in it, click on it, and the hyperlink delivers us to a virtual place, where we find a blue thing again that we can also click on. Hungarian humorist Frigyes Karinthy's *Automatic Device* is 'a simple apparatus where if you press a button at the top, another button pops out at the side, which you can press again.'

While there is not very much reliable and well-organized material, there is a lot more useless rubbish, or what we find is semi-useless, because we do not actually know how reliable it is, and the whole thing is truly messy. The Internet is not responsible for selection, in the same way as the printing machine was not either. Still, I think we are currently experiencing the age of new humanism. Similarly to the manner how in the fifteenth and sixteenth century with the help of the humanists the culture based on printing took the place of the culture of manuscripts, today we are witnessing how the virtual book is taking the place of the printed one. Literary studies may be at the cutting edge of the process. It is the task of just a few generations to transcribe the treasures of our civilization into the new code. The desire for content makes the expertise of those more valuable who know how and what to publish on the net. The philologist is now fulfilling the role the biologist once fulfilled, or the physicist did before that. Money is not a problem. Textbooks account for more than one-third of the Hungarian book market. Now even these are being replaced by on-line editions. In 1998, the government spent a great deal of money on the hardware, and on on-line textbooks within the framework of a so-called 'Sulinet' or to put it in English, the 'Schoolnet' program. The title of one of the columns on the *Wired* magazine is *Renaissance 2.0*, which suggests that they think the age of the information technology revolution should be regarded as the second, improved version of the Renaissance. The philologist was a characteristic figure of the Renaissance, who, having compared the variant readings of the

old codices, produced the new and reliable edition of the classical writer. Editing classics for on-line publication is still an equally important task, since not only researchers, but also higher- and intermediate education require the on-line accessibility of highly reliable texts.

This is the aim of the branch of science I, in need of a better term, would call information science literary studies. Information science literary studies provide the theory for the content industry.

2. Practice: f-book versus e-book

Now that we have to invent the book of the future, it is worth observing how difficult it was to invent the printed book during the Renaissance, when the previous information boom took place. It actually took several more years after Gutenberg for the first modern printed books to be born.

Even real experts find it hard to distinguish at sight between the early prints - the incunabula - and the codices. The initials in them were hand-painted, and the pages were often illuminated. Like in the age of copying, ligatures and abbreviations were still in use, which made the work of the typesetter difficult. It also took long decades for the sumptuous tradition of the title page to spread, since up until then the title was immediately followed by the text. It took long for more convenient sizes and materials to gradually become widely used (e.g. the use of wood for the binding boards was not renounced for a long time).

Publishers today are in a similar situation and, like their colleagues in the old days, are characterized by the lack of originality and creativity. I think the commercial failure of the electronic book may be due to the fact that it is no more than a poor attempt to imitate printed materials, in exactly the same manner as early prints were trying to imitate codices. It took almost half a century for the printed book to take its current shape, and we have no reason to believe that the invention of the on-line virtual book will happen at a faster pace. It is too early yet to attempt the standardization of the on-line book. Instead of e-books, we need f-books, where 'f' stands not only for free of charge but also for formal freedom.¹

¹ <http://www.f-book.com>

It is not only the form but also the content of e-books that is unsatisfactory. At time of the previous information boom humanists defined as their most important task the reliable publication of classics. The equivalent of this kind of book today could be the on-line critical edition, but these are rarely being produced. Publishers still have reservations about the net, they will go no further than producing CDs as a carrier for critical editions. As a result, public money for digitization ends up in the libraries rather than with the publishers. However, librarians' main duty is not to produce new books, but rather to safeguard and circulate them. Since publishers have not found their genuine niche on the net so far, professional philologists ever get commissioned to do on-line editing. The international collaboration for the digitization of codices shows signs of a certain barbarism: what I mean is that the pages of most books are digitized as a kind of picture, without recording the actual text. Numismatists rate highly those "barbaric coins" from the time of the fall of the Roman Empire on which the head of the unnamed ruler is surrounded by some decoration reminiscent of letters. Without the transcription of texts, images of codex pages are almost useless, since even those who have studied Latin for years and are relatively good at translation are unable even to copy even the most beautifully written codex page. I will elaborate on this issue later on, and will try to explain a more problematic regularity, namely that the more modern digital sign carriers are, the shorter their life expectancy is. The digitized images of manuscripts will have long been ruined when codices are still almost as beautiful as they are today.

Our on-line publications are produced as part of a course at the "Information Science Literary Studies Program" at Eötvös Loránd University in Budapest. The work is carried out by Gépeskönyv Ltd., a publishing company set up and owned by my students, which specializes in publishing and freely distributing virtual books of high cultural value, and which, surprise surprise, is a profitable enterprise. Like the humanists of old times, what we want to do is to keep track of and publish a critical edition of Hungarian classical texts.

In 1976 we started to build up a full repertory of early Hungarian poetry on the basis of multiple criteria, called *Répertoire de la poésie hongroise ancienne* (RPHA). The off-line

version of the database was published in 1992 in Paris, the on-line versions in 1993, and subsequently in 1999. Unlike the metric repertories published in books (Frank, 1966; Antonelli, 1984), and the repertories that use computers in an inefficient manner (Touber, 1975), this tool can be searched on the basis of heterogeneous criteria, can be programmed flexibly, and can be accessed free of charge on the net. And this was seen as cutting edge technology at the time of its creation.

Our first on-line critical textual publication was probably the first of its kind worldwide. At first we published only the most important source of the poems of the greatest poet of the Hungarian Renaissance, Bálint Balassi (1554-1594). In the so-called *Balassa Codex*, published in 1993, we replicated the text in a letter-perfect transcription, provided high-resolution facsimile copies complemented with rich notation and referencing. Later, in 1998 we published Balassi's whole poetic oeuvre. This publication provides the full text of every valuable handwritten and printed textual source in letter-perfect transcriptions and is accompanied with high definition photographs, providing the latest critically established text, apparatus and explanations.

Our next project was an experimental on-line edition in 1999, containing all the essays and articles of the most popular twentieth century Hungarian poet, Attila József (1905-1937).

The technological background of these publications is now outdated, therefore they are currently being reprogrammed. Five or six years is a long time in the history of the Internet. Our work, however, has brought some considerable achievements in literary history. In the Balassi edition, we reconstructed the original order of the poems in the poet's book of verse, and we were the first to discover the original title of the collection. The Attila József publication has succeeded in adding to the poet's oeuvre some 50 manuscript fragments that had so far not been known.

In 1999, at the Frankfurt Book Fair we had the opportunity to exhibit these theoretically new products: the on-line poetic repertory, and the two on-line critical editions. However, our other exhibit, the Kuhlmann-machine was a tremendous success (Teslár, 2000).

In 1671 Quirinus Kuhlmann, a German Baroque poet, published his famous poem (which he called a sonnet) in the volume titled *Himmliche Liebesküsse*. It is unlikely that he actually made a machine of this type, but he certainly did produce the algorithm for writing poetry. Unfortunately, it allows only the words in the middle of the lines to be changed, but even this provides for a great number of variations. Kuhlmann (and similar poets in the seventeenth century, from Georg Harsdörffer to John Peter) are clearly very popular nowadays. We can even find programs using the Kuhlman-algorithm on the net. However, I asked my students P. Benits and Á. Teslár (*ibid.*) to design a wooden machine that would function according to seventeenth-century logic. With the help of a brilliant cabinet-maker they managed to create the model.

This algorithm is important for us, as our motivation for looking to the past is not only for examples of failures (like the lack of originality in both early prints and e-books), but also to draw inspiration from it. Kuhlman is a good example for us.

Kuhlmann's poem is a combinatorial text, and combinatorics plays an important role in our publications, too. When compiling the inventory of ancient Hungarian poems, we considered every variation of a poem equal, as if we had been dealing with folklore rather than literature, and, as we shall see, we also used combinatorics in the critical apparatus of the Balassi publication. In literature we are in close agreement with Jorge Luis Borges, Raymond Queneau, Italo Calvino, the OuLiPo circle, who regard the Kuhlmann machine as a kind of *plagiat par anticipation*, 'anticipated plagiarism'.

3. Theory: textual studies

Let us define a literary text as a sequence of linguistic signs identified as a literary work in the literary institution.

Positivists treat the text as an object and wish to free it of the traces of earlier interpretations, of textual corruption. Following Lachmann's procedure, they would attempt to theorize textual derivation with the help of common mistakes found in the sources, and would determine the only authentic version of the text to meet the the author's intentions. Textual derivation is represented by a graph with no convergence, but this is in contradiction with experience. Computerized textual criticism makes sometimes the same mistake.

Theorists of hermeneutics believe that a text can only be known through individual readings, therefore they do not deal with unread texts. For them, the text cannot be separated from the act and agent of interpretation: we can only understand something that can understand us. Texts can only be interpreted in their relation to other texts. The boundaries of the textual universe are identical with the boundaries of the imaginable universe, since we cannot imagine anything that cannot be put into words. The precedence of thought by language implies that 'any comprehensible existence is language,' (Gadamer, 1986: 478) or that 'there is only text.' (Derrida, 1967: 227). The claim that reality outside language is epistemologically inaccessible implies for some that reality itself is of a linguistic nature. However, this statement can only be made from a position outside language, with a perspective on extra-linguistic reality, and by someone who is able to experience that which does not exist.

If the various interpretations of a text differ from each other, then the text itself remains self-identical. For an eclectic, new structuralism the self-identity of a text is questionable. Genetic textual criticism reveals the various strata of formulations on the basis of the author's manuscripts. Electronic text records prove that texts that exist in various forms do not possess any permanent qualities. In programmed on-line editions the selection among equal variations is made at random.

The faltering of the self-identity of the text has been observed by

many, but first of all by Bernard Cerquiglini (1989) in France. In Hungary we have made the additional observations which follow.

3.1. *Every text is plural*

The database of the RPHA registers all solid bibliographical, literary historical, and poetic information about each Hungarian poem written before 1601. The team of the RPHA worked with a pluralistic definition of the text. Accordingly, every feature of the text (e.g. its author, extent, literary model, genre, metrical structure, date of acquisition, dedication etc.) is a variant, and acquires different values depending on its occurrence in the various printed and manuscript sources. The text has no permanent features, the difference between high literature and folk art is only a difference in degree. The relational data model of the RPHA team allows the text to be linked to several (manuscript or printed) books, or a book to be linked to several texts. The record is neither the book, as would be in library databases, nor the text, as in literary studies, but a singular meeting point of a book and a text. This has led us to the realization that texts are always plural.

This new kind of text registration and critical edition offers the reader every textual stage of the history of the text as equivalent, since every textual event (every copying or publication) is a reading event. Reading event means a meeting of a text and a reader. It can even be regarded (since we are referring to the scribe, the typesetter and the publisher) as their interaction in the strictest sense of the word, as the fusion of their horizons (Gadamer, 1986: 311), which has created material traces.

3.2. *Contingent reading*

On the basis of the pluralistic conception of the text the RPHA proposed, in the first on-line critical edition we formulated the theory of programmed, combinatorial reading, and we named it 'stochastic reading' or 'random reading.' In those - not exceptional - cases where philologists cannot decide which branch of text derivation, which version to select, in the programmed critical edition a random generator - in more complicated cases a weighted random generator - starts operating. This method helped us to produce a

critical edition which provides the most reliable text possible, as well as to satisfy the current expectation to present a text in its historicity, as the product of an interaction between text and reader, as tradition.

We are working on a similar innovation in the process of transforming our second on-line critical edition: we are adding an option to make the programmed *apparatus criticus* invisible. The most problematic part of reading genetic publications is their extensive notation. Now there is an option to make it disappear at a click, and there remains only the selected stratum of the work.

3.3. *Text is something that can be copied accurately*

This definition is based on the experience of the RPHA team, namely on the realization that texts exist only as variations.

Whatever may be copied accurately is a text. Any part of the text that cannot be copied accurately is not part of the text.

Every database is textual; including those that contain visible or audible material. Digitization is the process of transforming into text something that was not text. Texts cannot be digitized. The digitization of texts means no more than the act of accurate copying, but with a different sign system.²

If we conceptualize the text as something that can be copied accurately, then digitization, this popular procedure of our age, has not really caused an essential shift. Neither did writing, the greatest invention of antiquity, bring about an essential change in its relation to speech, since the phoneme inherently contained the possibility of writing.³

If we define the text as something that can be replicated accurately, then it can be treated as a durable category from oral utterances to the Internet. This steadiness provides the possibility to talk about

² This definition is clearly identical with the definition of the text given by Srejder (1975): 'one container set is mapped into an alphabetical set which contains all the signs that can be replicated without limit. It is clear that object *a* cannot be regarded as the accurate copy of object *b*, if they are not identical with each other (accuracy rule), but if they are identical, then one cannot be the accurate copy of the other (replication rule). Therefore an object may not, but a relation may be accurately replicated. If in *M* alphabetical set object *a* and object *b* are both represented by unit *m*, then the relations $a \rightarrow m$ and $b \rightarrow m$ are equivalent'.

³ The above proves the main argument of Derrida (1967).

the universal history of texts, moreover to discover regularities in this history.

Although according to the definition we have applied, genetic codes qualify as texts, the historical rule does not apply to them.

3.4. While accessibility of texts increases, the life expectancy of text carriers is reduced

This phenomenon can be explained by the fact that accessibility is achieved by multiplication. The master copy (like the genetic information) only has to last as long as it is copied. The less energy expenditure copying requires, the more likely it is to occur. It is easier to replicate an e-mail than an inscription on a rock.

In biology it is exactly the least developed species that have perfected the multiplication process. A bacterium needs less energy to replicate its genetic code than an elephant, therefore it is more likely to happen.

3.5. The biology of text deterioration: literary history

The advantage of text deterioration is that the likelihood of the modified version to survive changes. It is the adaptation to the environment that textual critics call *lectio faciliior*, the disappearance of obscure passages. After a certain number of text modifications a new text (a new species) appears. Textual criticism prevents texts from adapting to their environment and thus surviving in a modified form.

3.6. High culture and popular culture will be replaced by free and paying culture

We are used to on-line libraries being free of charge, but recently even some on-line publishers have started placing their products at the users' disposal for free. The work of the publisher is being paid for, but it is not the consumer who pays. There is a similar phenomenon in software business, the source code of valuable program packages is made public, and their replication is not forbidden. There might be a similar turn taking place in the publication of scholarly books. There is an increasing number of publications

whose financing and consumption are not linked to each other, as it is in bookshops, but is separated, as in libraries. A movement is being organized against scholarly journals that are not prepared to make at least their previous issues freely available on the net. This change will hardly be limited to professional books and journals. It is likely that the present system of national (or supranational, European Union-based) educational subsidies will have to be restructured.

My assumption is that public money should be used for purchasing only things taxpayers can benefit from. Expenditure on education is a good example, since theoretically every scientific achievement serves common needs, and every statue may be seen by the public. Due to the information technology boom, taxpayers can actually have access to more and more educational achievements made with public money. But we can hardly expect the government or the European Union to ensure that all publicly funded educational achievements will reach each and every single taxpayer. However, if there is somebody who is prepared either for money or for free to get them to the taxpayers, they should not be prevented from doing so. Consequently, no institution should have the right to employ an entrepreneur who demands exclusivity to act as a mediator. Institutions' agreements that exclude third parties from educational mediation violate taxpayers' rights to accessibility.

Taxpayers' right to public collections must not be violated either, even if part of the collection was not actually created from their tax proceeds. This is why public collections should have no right to hire an entrepreneur either who stipulates exclusivity as a demand. Public collections' agreements that exclude third parties also violate taxpayers' right to accessibility. For instance, those public collections which sold the exclusive right for digital mediation of works of art they were looking after to a well-known software manufacturer, made an agreement which breaches the law, which is immoral, and therefore invalid. Public collections do not have unlimited rights concerning the possession of objects they have been entrusted with.

At the same time, the producer of a work of art can accept money for the mediation rights of his/her work, even if it was subsidized with public money. For instance, for the interest of its

members a subsidized music group may enter into an agreement with any broadcasting company or publisher. Subsidized scientists, unless restricted by an agreement with their employer, may sell the patent of their intellectual property. However, public money should not be spent on work whose producers demand money for its mediation rights, and if it is not provided, their work does not become accessible to taxpayers. Public money should not be given to scientists who do not make available their publications immediately and free of charge to the taxpayers. For instance, we should never erect a public statue or build a public building that may be viewed only for a fee. Let us not contribute public money to the royalty of such literary work that may not become freely available on the shelves of virtual libraries. Publicly maintained universities and research institutions should not employ scientists if their achievements - with the exception of defence research - may not become immediately available to the public on the net.

On the other hand, we must not mediate the work of those authors, unless their permission is granted, who did not use public money in the process. For instance, the use of the software that makes possible the on-line exchange of musical data is only right, if the data were created with the help of public money, like in the case of symphonic orchestras. It is wrong to facilitate the unpaid use of services produced by taxpaying businesses (e. g. successful pop music groups.)

To sum up, it is the community who possesses the right to use all the data that were created from the expenditure of the community, rather than the institution where those data were created. Taxpayers are likely to want - on the basis of mutuality - other taxpaying communities to have access to the educational rights granted to them. This is why they will refuse insularity, whether at governmental or EU level. If following the mandatory, but fundamentally free school system and the public media broadcast, in other areas of culture too, costs will no longer be borne directly by the consumer, then culture and education will come closer to other large social providers like health care and pension provision. In this field, according to the centuries old tradition of state cultural financing, the European Union has the chance to achieve a competitive advantage over the United States, however not over the Asian region.

Notice

This article is an intellectual property and falls under the force of the GNU Public Licence (GPL, © 1989, 1991), therefore it is not separable from the following notice.

© Iván Horváth, 2004. Everybody has the right to fully or partially, by any means, copy this article either on paper or on another data carrier, to store it there without limit, to distribute it free of charge, or to sell it for money, provided that this copyright and the permitting note is indicated in a legible manner on every partial or full copy. It is not allowed to make modified versions without permission. Translation does not qualify as modification.

REFERENCES

- ANTONELLI R., *Repertorio metrico della scuola poetica siciliana*, Centro di Studi Filologici e Linguistici Siciliani, Palermo, 1984.
- BALASSI B., <http://magyar-irodalom.elte.hu/gepesk/bbom/>
- CERQUIGLINI B., *Éloge de la variante*, Éditions du Seuil, Paris, 1989.
- DERRIDA J., *De la grammatologie*, Éditions de Minuit, Paris, 1967.
- FRANK I., *Répertoire métrique de la poésie des troubadours*, I-II, Champion, Paris, 1966.
- GADAMER H.-G., *Wahrheit und Methode*, J. C. B. Mohr, Tübingen, 1986 (*Gesammelte Werke*, I.)
- HORVÁTH I. (ed.), H. HUBERT G., FONT ZS., HERNER J., SZÖNYI E., VADAI I., GÁL Gy., *Répertoire de la poésie hongroise ancienne. Manuel de correction d'erreurs dans la base de données*, I-II, Éditions du Nouvel Objet, Paris, 1992.
- HORVÁTH I. (ed.), H. HUBERT G., FONT ZS., HERNER J., SZÖNYI E., VADAI I., GÁL Gy., KARSAI R., <http://magyar-irodalom.elte.hu/repertorium/>
- JÓZSEF A., <http://magyar-irodalom.elte.hu/ja/>
- SREJDER J. M., *Egyenlőség, hasonlóság, rendezés* (translated by A. Vargha), Gondolat, Budapest, 1975.
- TESLÁR Á., <http://magyar-irodalom.elte.hu/colloquia/000601/teslar.htm>
- TOUBER A. H., *Deutsche Strophenformen des Mittelalters*, Metzler, Stuttgart, 1975.