

Text Comparison and Digital Creativity An Introduction to the Colloquium Theme

Wido van Peursen and Ernst Thoutenhoofd

(Underlined names refer to contributions to the colloquium; quotations come from the abstracts, which can be found at www.knaw.nl/agenda/pdf/20081030_VKS_programme.pdf)

1. Text Comparison and Digital Creativity

Text comparison is at the very heart of philology. 'For research in textual composition we need instruments that allow us to compare various literary texts: concordances that present words in context and synopses that present parallel texts.' (Talstra) The comparison may involve various sections of the same text (cf. Talstra's reference to concordances), texts that in some way are 'parallel' (cf. Talstra's reference to synopses), and texts that are more loosely related, containing, for example motifs and themes that frequently reoccur in certain cultures, or other intertextual relationships. Parallel texts, texts that in one way or another are witnesses to the same composition, can be described as a continuum that 'ranges from the extreme of extensive verbatim quotation, on the one hand, to the point where no relationship is discernible, on the other'.¹ The philological activity of text comparison goes back to Antiquity, and is attested, for example, in Origen's Hexapla (Weinberg).

'Digital creativity' is at first sight a paradox. Doesn't 'digital' refer to calculation and sorting of data, whereas 'creativity' refers to everything that escapes calculation: subjectivity, the master's eye, and the intuition of the experienced scholar? Applications of ICT in, for example, bio-informatics, in which the computer interacts with the human researcher to distract knowledge from immense data collections, or studies in artificial creativity, show that this is not the case. Definitions of creativity often include the notions of plurality and variation. Working with huge amount of data requires creativity to handle them, even, or perhaps even more, when they are handled in interaction with ICT. 'Digital creativity' is 'the ingenuity of the human beings who created algorithms for processing language' (Weinberg).

The title of the colloquium brings 'text comparison' and 'digital creativity' together. Its aim is to explore recent developments in comparative text scholarship brought about by the use of ICT. The papers have been organized around four themes: 'Texts as artefacts', 'Texts as objects of transmission', 'Texts, knowledge creation and representation' and 'Continuation and innovation in digital text scholarship'.

2. Presence and Meaning

'Presence' and 'meaning' figure in the subtitle of the colloquium. They refer to the role of materiality ('physics') and meaning ('meta-physics') in text scholarship, and they are related to two of our colloquium themes, namely 'Texts as artefacts' and 'Texts as objects of transmission'.

Traditionally text scholarship is concerned with interpretation, the attribution of meaning. It is related to metaphysics, which Hans Ulrich Gumbrecht describes as 'an attitude that gives a higher value to the meaning of phenomena than to their material presence', and that points to a worldview that always wants to go 'beyond' what is physical.

'Presence' refers to the materiality of things, to the physical aspects, to *What Meaning Cannot Convey*.² According to Gumbrecht, Descartes' dichotomization between 'spiritual' and 'material' led to the introduction of the subject/object paradigm, in which 'to interpret the world means to go beyond its

¹ M.O. Wise, *A Critical Study of the Temple Scroll from Qumran Cave 11* (Studies in Ancient Oriental Civilization 49: Chicago/Illinois, 1990), 207; cf. W.Th. van Peursen and E. Talstra, 'Computer-Assisted Analysis of Parallel Texts in the Bible. The Case of 2 Kings xviii-xix and its Parallels in Isaiah and Chronicles', *Vetus Testamentum* 57 (2007) 45-72.

² Part of the title of Gumbrecht's *The Production of Presence. What Meaning Cannot Convey*.



material surface or to penetrate that surface in order to identify a meaning (i.e., something spiritual) that is supposed to lie behind or beneath it'.³ He describes this development as follows:

Medieval Christian culture was centered on the collective belief in the possibility of God's real presence among humans and in several rituals, most prominently the Mass, that were meant to constantly produce and renew such real presence. (...) In modern culture, in contrast, beginning with the Renaissance, representation prevails over the desire for real presence. Representation is not the act that makes 'present again', but: those cultural practices and techniques that replace through an often complex signifier (and make thus available) as 'reference' what is not present in space or time. (...) My innovative thesis lies in the claim that, ever since the historical moment that we call the 'crisis of representation', around 1800, our cultures has developed a renewed longing for real presence.⁴

Although Gumbrecht sketches a long pre-history of the re-awakened interest in the 'presence', we think that it is only in the second half of the last century that the hegemony of meaning is seriously threatened. Interpretation, the academic, analytical and intellectual activity that concerns the attribution of meaning, has come under attack. In 1966 Susan Sontag wrote an essay 'Against Interpretation',⁵ in which she calls interpretation 'the intellect's revenge upon art'. She finished with the words, 'in place of a hermeneutics we need an erotics of art'. Robert Alter speaks of 'the heresy of explanation'⁶ (Weinberg) and claims that 'the unacknowledged heresy underlying most modern English versions of the Bible is the use of translation as a vehicle for *explaining* the Bible instead of representing it in another language',⁷ the result is that 'the modern English versions—especially in their treatment of Hebrew narrative prose—have placed readers at a grotesque distance from the distinctive literary experience of the Bible in its original language'.⁸ George Steiner speaks of the 'Byzantine dominion of secondary and parasitic discourse over immediacy, of the critical over the creative'.⁹ And Gumbrecht 'challenges the broadly institutionalized tradition according to which interpretation is the core practice, the exclusive core practice indeed, of the humanities'.¹⁰

The 're-awakened interest in the presence' is visible in the attention paid to texts as artefacts and the material aspects of the carriers of texts, which hardly receive any attention in traditional text scholarship. In the field of biblical scholarship, for example, most standard works on textual criticism hardly pay any attention to mise-en-page, delimitation markers, paleography, and codicology. In the last decades this situation has changed and various research groups dealing with these aspects of textual transmission have been established, such as the Pericope Research Group (see www.pericope.net).

Dissatisfaction with the analytical academic attribution of meaning has led to suggestions to avoid the distance-creating interpretation and to strive for immediacy. However, opinions differ about what this immediacy is. For Gumbrecht it is the material, physical presence of the objects of the world. For Alter it is the appreciation of the artistic value of the literary work, including all its ambiguities that philologists try to explain away because of their strife for clarity. Steiner says that his desire for 'a greater closeness to the things of the world and for a greater intensity in this contact than our everyday worlds would allow for' is 'transcendental'.¹¹ Readers of the Bible can find an excuse for avoiding the

³ Gumbrecht, *Production of Presence*, 26.

⁴ Gumbrecht, *Powers of Philology*, 11.

⁵ Cf. Gumbrecht, *Production of Presence*, 10.

⁶ Robert Alter, *The Five Books of Moses. A Translation with Commentary* (New York: Norton, 2004), xvi.

⁷ Alter, *ibid.*, xix.

⁸ Alter, *ibid.*, xvii.

⁹ Steiner, *Real Presences*, 38.

¹⁰ Gumbrecht, *Production of Presence*, 1.

¹¹ Steiner, *Real Presences*, 146.



analytical interpretation and looking for direct, spiritual access to the biblical text in the century-old tradition of the *lectio divina*.¹²

Both Gumbrecht and Steiner advocate 'Real Presences'. Both will agree that going to a concert where Händel's *Messiah* is performed gives a better access to that work than any musicological analysis of it, and that visiting the National Gallery of Berlin to see Casper David Friedrich's *Monk by the Sea* gives a more direct access to this painting than reading an interpretation of it by an art historian. When, however, it comes to texts, the differences between their views come to the front. For Gumbrecht, 'presence' refers to the spatial relationship to the world and its objects and hence to the material carriers of texts. In Steiner's view the effect of the confrontation with the primary (e.g. a literary text) is twofold: It gives direct access both to the physical material, avoiding the detour of interpretation, and to that which goes beyond description and interpretation, namely the transcendental real presence.

How do 'presence' and 'meaning' relate to digital text scholarship? On the one hand, digital text scholarship has to face the challenges to interpretation put forward by Steiner and others. According to Steiner, 'what looks to be certain is that the criteria and practices of quantification, of symbolic coding and formalization which are the life breath of the theoretical do not, cannot pertain to the interpretation and assessment of either literature or the arts.'¹³ This claim provides a challenge to the analytical and mathematical approaches that characterize the computational analysis of texts, even more than to traditional philology. It raises the question how the computational, analytical work done in digital scholarship relates to the subjective moods of interpretation and intuition that characterize traditional philology (cf. below, 'Scholarly and Scientific Research').

On the other hand, if our claim that 'the renewed longing for real presence' increased considerably in the second half of the last century, is correct, this development overlaps with the technological innovations of the digital age, which made new realizations of 'presence' possible. The new technologies support the study of texts as artefacts (cf. the projects presented by [Parker](#), [Schmid](#), [Boyle](#), and [Zuckermann](#)), and help discover 'what stories the messenger can tell beyond the explicit message' ([Boyle](#)). The electronic presentation of textual artefacts raises the question as to the similarities and differences between the ancient production and its electronic reproduction ([Parker](#)).

3. Scientific and Scholarly Research

The main focus of the interest of the colloquium has been defined as follows:

The spread of digital technology across philology, linguistics and literary studies suggests that text scholarship itself is taking on a more laboratory-like image. The ability to sort, quantify, reproduce and report text through computation would seem to facilitate the exploration of text as another type of quantitative data (akin to protein structures or geographic features of the seabed). However, developing this potential also highlights text analysis and text interpretation as two increasingly separated sub-tasks in the study of texts. The implied dual nature of interpretation as the traditional, valued mode of *scholarly* text comparison, combined with an increasingly widespread reliance on digital text analysis as *scientific* mode of inquiry raises the question as to

¹² Cf. Reedijk, *Zuiver lezen. De Bijbel gelezen op de wijze van de vroegchristelijke woestijnvaders* (PhD dissertation Vrije Universiteit Amsterdam; Budel: Damon, 2006).

¹³ Steiner, *Real Presences*, 79. But Steiner does not completely reject the philological analysis. He refutes F.R. Leavis' claim that "linguistics has nothing to contribute to the understanding of literature", because, "An informed alertness to the phonetic, lexical, grammatical instrumentalities of a text both disciplines and enriches the quality of interpretative and critical response" (p. 84). "Therefore radical doubts, such as those of deconstruction and of the aesthetic of misreadings, are justified when they deny the possibility of a systematic, exhaustive hermeneutic, when they deny any arrival of interpretation at a stable, demonstrable singleness of meaning. But between this illusory absolute, this finality which would, in fact, negate the vital essence of freedom and the gratuitous play, itself despotic by its very arbitrariness, of interpretative non-sense, lies the rich, legitimate ground of the philological (pp. 164–165).



whether the reflexive concepts that are central to interpretation – individualism, subjectivity – are affected by the anonymised, normative assumptions implied by formal categorizations of text as digital data. This calls for a reconsideration of the scholarly/scientific and intellectual/computational ‘co-production’ of presence and meaning of text in philology.

The contributions to the colloquium indeed provide a number of examples in which textual analysis receives a more ‘scientific’ character, in which formalization is used as a means to overcome the individualism and subjectivity that characterizes much philological research, or in which generalization of analytical procedures and formal and systematic registration of the data help develop research strategies that allow for verification of its conclusions. Dahlström addresses the question whether editions are presentation of facts or are hermeneutical documents and subjective interpretations. Salemans shows how the computer can help implement an intersubjective, repeatable and controllable theory, and Lavagnino presents a systematic approach to emendations. All these contributors agree, however, that ‘scholarly judgment and experience are still needed’ (Lavagnino) and that the subjective or intersubjective human judgment is indispensable.

4. Continuation and Innovation

The basic question that arises when we discuss innovation brought about by ICT is, whether there is really methodological innovation: ‘Do we speed up classical techniques, or do we develop a new domain of techniques for access to the classical texts?’ Can we get beyond imitating the *classical instruments for sorting lexical materials* (concordance, lexicon) and for comparing parallel literary texts and fragments (synopsis)? (Talstra) Is our use of the digital medium in fact an imitation of the medium of the book, or do we start to employ the salient characteristics of the new medium? (Van der Weel; cf. below, section 5). What seems very innovative, like the use of the calculation and sorting force of the computer, has predecessors in the indexes and concordances of the Masorettes from long before the digital era (Weinberg).

Although all colloquium contributions in one way or another touch upon this question, we have organized one session that is more specifically devoted to the theme ‘Tradition and innovation in digital text scholarship’, to which we gave the following description:

Many electronic instruments for textual scholarship that have become available over the last decades are imitations of traditional tools. Instruments for word searches in digital texts, for example, fulfill the same function as traditional concordances. Where does the computer lead to new research strategies? How are reflexive concepts that are central to interpretation – individualism, subjectivity – affected by the anonymised, normative assumptions implied by formal categorisations of text as digital data? What can be the role of formalisation within a mainly interpretative framework of analysis?

The question as to the innovation brought about by ICT plays a role at various levels. Questions that can be asked are: How can the computer serve as a research instrument, thanks to its capability of calculation and sorting? And how do scholars use the electronic means of communication? How does the computer change our basic notions of language, text, authorship, and others?

The innovation brought about by ICT changes the basic concepts of philology. The notion of ‘text’ as a means of communication changes due to Electronically Mediated Communication, which differs both from speech (e.g. tone-voice substitutions) and from writing (e.g. hypertext linkage, framing) (Crystal; see also below, section 5). The notion of ‘text’ as object of transmission has changed due to the potential of Electronic Editing. Traditionally scholarly text editions contain a main text and a critical apparatus, and the establishing of the main text and the selection of variants are based on the text-critical research of the editor. However, the conceptualization of the notions of ‘text’, ‘variant’, and



‘edition’ should be redefined in digital text scholarship (Dählström; Lavagnino).¹⁴ Especially in the study of interrelated textual artefacts from before the invention of printing the new models for the representation of textual variation are useful. ‘The real challenge is the fact that biblical texts are part of the large collections of classical texts that have been produced long before the invention of the art of printing (...) We only possess individual copies of a text, none of them being identical to any other existing copy’ (Talstra).

Not only the practice of text editing, but also that of text comparison changes. Processes of text production and transmission have led to a multitude of parallel compositions.¹⁵ The traditional ways of dealing with these works in synopses or critical text editions are bound to the limitations of paper format and size. In the last decades we have seen a number of initiatives to use the computer for text representation and comparison. An interesting example is the edition of Ogdoas Scholastica, which implements text as hypertext. According to Øhrstrøm & Peterson, ‘understanding of Lorhard’s book can be greatly facilitated by the use of various kinds of digital creativity’.

In this context we should also note the interaction between technological innovation and the ‘re-awakened desire for presence’ mentioned above, at the end of section 2.

5. Knowledge Creation and Representation

Finally, we want to take one further step in the reflection on basic concepts, data analysis, and information retrieval, in the session ‘Knowledge creation and representation’, to which we gave the following description:

Text, as a record of ideas, a means to construct author(ity), and material carrier of communication between humans has been central not only to philology, but to scholarship generally. Text is knowledge represented as matter: visible and revisitable, portable and measurable. As discipline focused on understanding texts and change in texts critically, philology is therefore a unique scholarly resource for understanding ways in which text alters under conditions of new technology, but of course knowledge of text in philology is itself also a text, both epistemologically specific and formally encoded (theorised). Does new technology make philological approaches and insights into the nature of text more transparent for other scholars? What broader challenges, shared interests and opportunities emerge as text comparison becomes part of a wider move towards integrated forms of (collaborative) e-research and the multi-purposing of data collections?

The movement to a more laboratory-like image of digital scholarship (see above, section 3) requires from philologists that they start with the rough data, such as the sequences of graphemes that constitute a text. When the data are processed to be useful, and meaningful connections between data are established, so that they provide answers to ‘who’, ‘what’, ‘where’, and ‘when’ questions, the data become information. The appropriate collection of information results in knowledge, which provides questions to the ‘how’ questions.¹⁶

¹⁴ See also D. Buzzetti, ‘Digital Representation and Text Model’, *New Literary History* 33 (2002) 61–88. P. Eggert, ‘Text-encoding, Theories of the Text, and the “Work-Site”’, *Literary and Linguistic Computing* 20 (2005) 425–535; D. Schmidt, ‘Graphical Editor for Manuscripts’, *Literary and Linguistic Computing* 21 (2006) 341–351.

¹⁵ On ‘works’ (rather than ‘texts’) that are subjected to ongoing processes of composition and revision, see Hans Voorbij, ‘The *Chronicon* of Helinand of Froidmont: A Printed Edition in an Electronic Environment’, in H.T.M. van Vliet, *Produktion und Context* (Beihefte zu Editio 13; Tübingen 1999), 3–12, esp. 5: ‘The editorial team decided to produce a “historical-critical, genetic edition”. This type of edition tries to give an understanding of the process of composition and textualization of the *Chronicon*. It is not “the” text of the *Chronicon*, nor even “a” text thereof that will be edited, but rather the *Chronicon* as a “work”.’

¹⁶ For the definitions of ‘data’, ‘information’ and ‘knowledge’, we have used www.systems-thinking.org/dikw/dikw.htm.



What is the role of the computer in this process of knowledge creation? What is the role of the analytical potential and the calculation power of the computer in the analysis of the data, the retrieval of information, and the creation of knowledge? So many data can now be made easily available, but how to develop abstract concepts for the preservation of ‘concrete information’? ([Zuckermann](#)) Does the digital medium lead to new ways of collaborative knowledge creation? And what role does the computer play in knowledge representation? As noted in the preceding section, Electronically Mediated Communication differs from speech as well as writing and leads to new notion of text ([Crystal](#)). Hyperlinking, for example, can drastically change the representation of information ([Øhrstrøm & Peterson](#)). The internet provides possibilities of worldwide data distribution, but it provides new challenges regarding storing, distribution and presentation of information ([Zuckermann](#)), and new questions concerning responsibility and authorship ([Crystal](#)) and copyright ([Schmid](#)). New tools for collaboration and text annotation raise the question as to how we can overcome the individualism of humanities scholars ([Zafrin](#)). Interesting case studies concerns scholars annotating 14th century Italian texts on the site of the Virtual Humanities Lab, Brown University ([Zafrin](#)) and the digital workflow of the Virtual Manuscript Room ([Schmid](#)), which employ of the advantages of the electronic medium for collaboration.

However interesting and innovative all these applications of the digital medium are, we have to be aware that we are still at the very first beginning of the digital era and that, as [Van der Weel](#) puts it, we still live in ‘the Order of the Book’ rather than the digital order. [Van der Weel](#) recalls that in the past the new means of communication had a long way to go before their salient characteristics were fully employed. Their usage started as imitation of the familiar forms of communication: writing started as speech written down and printed books imitated manuscripts. So the challenge is how we can make full use of the salient characteristics of the unfamiliar new medium and what is the best possible use we can make of it for the task of creating ‘knowledge instruments’ such as thematically structured research collections. In order to answer this question, [Van der Weel](#) explores the nature of the digital medium.

