

Gerrit Brüning, Katrin Henzel and Dietmar Pravida

## Multiple Encoding in Genetic Editions: The Case of "Faust"

---

### Warning

The contents of this site is subject to the French law on intellectual property and is the exclusive property of the publisher.

The works on this site can be accessed and reproduced on paper or digital media, provided that they are strictly used for personal, scientific or educational purposes excluding any commercial exploitation. Reproduction must necessarily mention the editor, the journal name, the author and the document reference.

Any other reproduction is strictly forbidden without permission of the publisher, except in cases provided by legislation in force in France.

**revues.org**

Revues.org is a platform for journals in the humanites and social sciences run by the CLEO, Centre for open electronic publishing (CNRS, EHESS, UP, UAPV).

---

### Electronic reference

Gerrit Brüning, Katrin Henzel and Dietmar Pravida, « Multiple Encoding in Genetic Editions: The Case of "Faust" », *Journal of the Text Encoding Initiative* [Online], Issue 4 | March 2013, Online since 08 March 2013, connection on 03 June 2013. URL : <http://jtei.revues.org/697> ; DOI : 10.4000/jtei.697

Publisher: Text Encoding Initiative Consortium

<http://jtei.revues.org>

<http://www.revues.org>

Document available online on:

<http://jtei.revues.org/697>

Document automatically generated on 03 June 2013.

TEI Consortium 2013 (Creative Commons Attribution-NoDerivs 3.0 Unported License)

Gerrit Brüning, Katrin Henzel and Dietmar Pravida

# Multiple Encoding in Genetic Editions: The Case of "Faust"

## 1. Introduction

- 1 Since the 1950s, editorial theory and practice in the English-speaking world, France, and Germany have increasingly addressed issues of literary genesis: drafts, versions, the writing process, and methods of editorial rendering. While these particular scholarly traditions are quite distinct and will continue to be so for some time to come (see Van Hulle 2004 for a comparative assessment of these traditions), the development of common encoding standards for digital editions supports the formation of a zone of convergence for different traditions. The new chapter 11 of version 2.0.0 of TEI P5 is probably a first step in that direction (TEI Consortium 2011a; cf. the discussion in Brüning, Henzel, and Pravida 2012).
- 2 The new edition of Goethe's *Faust* consists of four basic parts:
  1. a digital archive
  2. a digital genetic edition
  3. a critically established text of *Faust* (both digital and print)
  4. a printed facsimile edition of selected manuscripts<sup>1</sup>
- 3 The digital archive contains facsimiles, transcriptions, and metadata. On this basis, the genetic edition attempts to reconstruct and visualize the genetic relationships within and between the documents. Diplomatic renderings, as well as presentations of the texts and their variants in different forms of critical apparatus, are to be generated automatically from the basic transcripts. A critical text of *Faust* will be established by taking into account the complete record of variants. The edition combines two different aims: to meet the standards established by printed scholarly editions and by the tradition of editorial theory, and to adopt recent developments in the field of electronic editions, thereby exploring the evolving possibilities of the digital medium.
- 4 The project's principles of encoding were developed in conjunction with the workgroup (part of the TEI SIG on Manuscripts) that developed the "Encoding Model for Genetic Editions" (Workgroup on Genetic Editions 2010), which formed the basis for the revised chapter 11 of TEI P5. A rethinking of the basic tenets of TEI markup and of our way of adopting the TEI standard for our edition has led us to disentangle one single form of "text encoding" into two different but closely related forms of encoding in order to provide both a diplomatic rendering of the layout of each manuscript (generated from a "documentary" transcript) and a representation of textual genesis (generated from a "textual" transcript). This differentiation complies with the theoretical distinction between record and editorial interpretation as proposed by Hans Zeller (1995): the documentary transcript records inscriptional phenomena, whereas the textual transcript records the genesis of a text.
- 5 Multiple encoding of similar (if not the same) information has some obvious disadvantages. The disadvantages can, however, be dealt with by the use of a sufficiently powerful collation tool. Disentangling two perspectives does not imply that they are independent from one another. On the contrary, their close correlation is of great importance for the reconstruction of genetic relations.

## 2. Genetic Editing on a Large Scale: The Case of Goethe's *Faust*

- 6 Goethe's work on *Faust* lasted sixty years, from about 1775 until shortly before his death in 1832. Manuscript material comprising about 2,300 pages written by Goethe and his scribes is

extant, most of it dating from 1825 or later. Since then considerable effort has been devoted to the investigation of the material. But despite a number of very valuable special studies (the latest being Bohnenkamp 1994) the genetic process still has not been analyzed in its entirety. The only comprehensive account of the textual variants dates back to the nineteenth century (Goethe 1887–88).<sup>2</sup> The presentation of such a large body of manuscripts and of a wealth of scholarly knowledge accumulated within a century of assiduous research is a daunting task. The material and the corresponding information should be made easily accessible in a way that enables and fosters further research. Any editorial uncertainties are not to be ignored, and conflicting scholarly opinions have to be considered.

7 In the following, we will concentrate on the digital archive and the genetic edition, especially the underlying encoding model, which not only applies the set of tags available in the TEI Guidelines but also extends it at some points.

8 The *digital archive* is a research platform that makes accessible all manuscripts and the authoritative prints of *Faust*. Facsimiles, diplomatic rendering, and physical descriptions (metadata) of all witnesses will afford the user possibilities of in-depth manuscript research.<sup>3</sup> The *genetic edition* gives an account of the textual alterations within single manuscripts and of the stemmatic relationships obtaining between them. The analysis of both will be visualized using not only various forms of critical apparatus and text presentation but also charts and other graphical representations of data.

9 The diplomatic rendering, as well as the presentation of texts and their genesis, is each based on XML-encoded transcripts. Any piece of information to be conveyed to the user has to be put into these transcripts. Therefore, the richness and quality of the information hinge on the constraints imposed by formats and rules of encoding. The editors have adopted the principle that particular problems should be solved according to general and explicit rules of encoding. The encoding principles and strategies devised for the *Faust* edition should in principle be transferable to editions of other authors with comparable habits of writing. As will be shown, problems that appear to be particular may turn out to be closely related to very general aspects of textuality.

10 Throughout this paper, reference will be made to the following example from a revised autograph copy of *Faust II* (fig. 1), which poses a problem of encoding that may be quite typical for modern literary manuscripts:

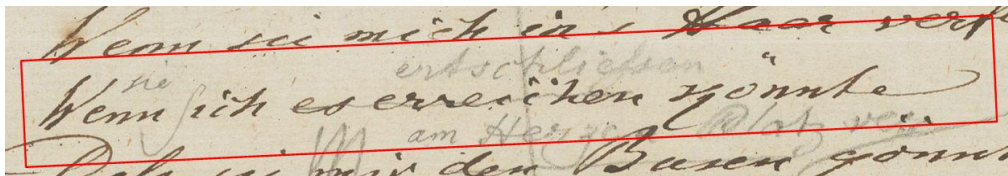


Figure 1: Detail from a revised autograph (Goethe and Schiller Archive Weimar, GSA 25/W 1427), *Faust II*, line 5142.

11 The line in the highlighted box of figure 1 reads "Wenn ich es erreichen könnte" ("If I could achieve" [my goal]), written with ink. When revising the manuscript with lead pencil, Goethe modified the text by putting "s" in front of "ich" ("I"), thereby changing the personal pronoun into a reflexive pronoun, struck out the words "es erreichen", and wrote "sie" ("she") and "entschließen" ("to bring [herself] to") above the line.

### 3. Genetic Encoding in TEI

12 The overall conception of an integrated electronic edition requires that all forms of presentation of information on manuscripts and texts be generated automatically from the transcript files. For this purpose, both inscriptional phenomena and textual alterations must be accounted for in detail. Version 1.3.0 of P5 offered a variety of tags and attributes to do this (TEI Consortium 2008, 72–78, 335–74). According to this version of P5, the example given in figure 1 may be encoded like this:

```
<l>Wenn <add place="above" hand="#g_bl_lat">sie</add>
  <add hand="#g_bl_lat">&#x17f;</add>ich
  <subst hand="#g_bl_lat">
    <del rend="strikethrough">es erreichen</del>
```

```

    <add place="above">entschlie&#x17f;sen</add>
  </subst> könnte
</l>

```

Example 1: Encoding of figure 1 in accordance to version 1.3.0 of P5.

- 13 Example 1 actually reflects the encoding practice of an early stage of the project. The structural divisions of the text and the textual alterations were recorded with the elements shown in example 1: <l>, <del>, <add>, and <subst>. Information on specific material properties—mode of deletion, lineation of the manuscript, scribal hand, writing medium, position of added words—was conveyed by the elements <hi>, <lb/>, and <handShift/>, and by the attributes @hand, @rend, and @place (for the values of the attribute @hand, see section 4.1 below). The transcript contains all information necessary for generating a diplomatic rendering since the approximate spatial position of the written characters can be inferred from it. Likewise, it is possible to deduce the primary and the final wording from the encoding so that some sort of text presentation and critical apparatus may be feasible. However, a closer look reveals that the encoding in example 1 is unsatisfactory in many respects.
- 14 The sequences "sie" and "entschließen" lack any explicit indication of their precise horizontal position. In the relatively simple case of example 1, it might still be possible to generate an adequate diplomatic rendering because the value "above" can be interpreted as "right above." In more complex cases (such as retraced letters or accumulated additions in more than one line or distributed over a line) the result would inevitably fall short of the standards set by printed editions (for example, because the values "margin", "top", and "bottom" for @place are far too vague to sufficiently indicate the spatial position of added words or passages). The set of tags offered in P5 version 1.3.0 is extensive, but it was not designed to fit all the needs of a full-fledged diplomatic edition.
- 15 Despite its lack of precision, the encoding in example 1 is too much focused on single aspects of the writing process to appropriately cover textual alterations. In the encoding, "sie" (she) appears to be merely added, whereas "ich" (I) is, as it were, supplemented by an "s." A thorough account of the textual alterations would have to indicate that one single substitution takes place: a personal pronoun ("ich"), a direct object ("es"), a verb ("erreichen"), and the first person singular conjunctive preterit of können ("könnte"—if *I* could) are jointly substituted by another pronoun ("sie"), a reflexive pronoun ("sich") together with the corresponding verb "entschließen", and a homograph third person form ("könnte"—if *she* could). In the encoding of example 1 there is not the faintest indication that all three individual changes made with pencil belong to one and the same textual alteration. The fact that the "ich" is substituted by "sie" is completely blurred by the only superficially adequate "<add>f</add>ich". The silent grammatical conversion of the finite verb is not expressed at all, as there is no tagging around the word "könnte" to mark this change.
- 16 In short, the initial way of encoding was neither rich enough to generate a reliable diplomatic rendering, nor did it meet the requirements of a *text* encoding in the proper sense of the word, i.e., an encoding that respects the linguistic nature of written texts. The need to extend the existing tagset according to genetic research interests has been felt in other editions and projects as well (Saller 2003; Vanhoutte 2002; Pierazzo 2009). To satisfy this need, the Workgroup on Genetic Editions, part of the TEI Manuscripts SIG, drafted an "Encoding Model for Genetic Editions" (Workgroup on Genetic Editions 2010). It has been largely incorporated in chapter 11 of version 2.0.0 and later of P5 (TEI Consortium 2011a).<sup>4</sup> It is not our aim to give an overview of all new elements and attributes introduced in the new version of chapter 11. Instead we will discuss below the application of the new tags to the example given in figure 1.
- 17 It is now possible to record the sediments of the writing process<sup>5</sup> as well as the textual alterations in a much more differentiated way. Temporal relations or syntactic dependencies of alterations can be expressed properly. As a consequence, the complexity of markup keeps increasing in the process of adapting encoding strategies to editorial demands. From a certain point on, the growth of the encoding complexity seriously impairs the readability of the encoding for humans and makes it hard for machines to process. However, there is still no guarantee that all indispensable information can be adequately stored in the transcript.

18 In fact, our project reached an impasse caused by bloated encoding even before the new tags were available. In particular, the encoding was impaired by an increasing number of conflicts, many of them caused by overlapping hierarchies (Renear, Mylonas, and Durand 1996). When it finally proved unavoidable, we decided to split the encoding into two distinct transcripts: a documentary transcript of each page of a manuscript and a textual transcript of the whole manuscript. That is to say, the encoding approach reflected in example 1 was abandoned in favor of a rigorous disentangling (fig. 2).

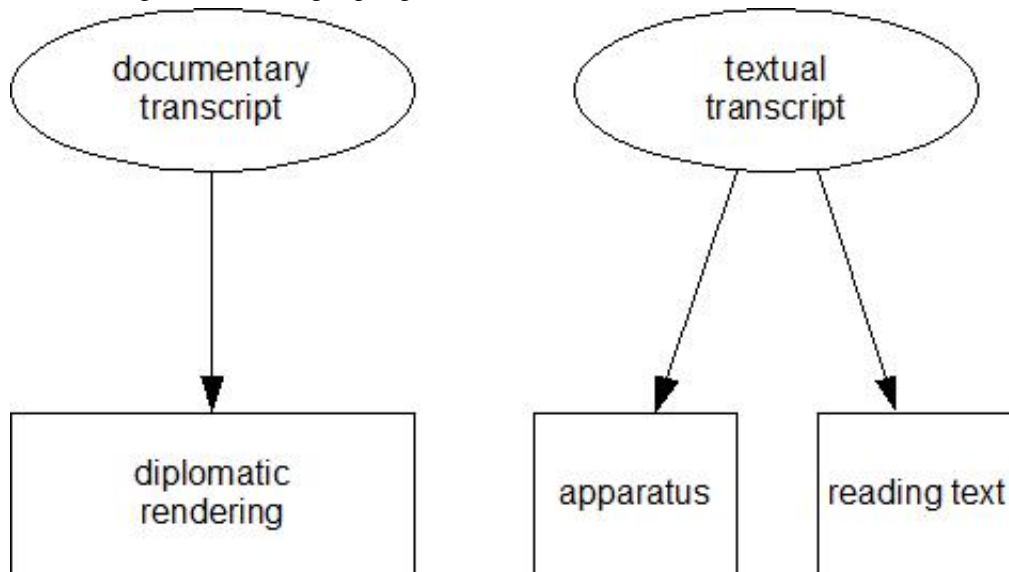


Figure 2: Encoding perspectives / derivation relationships / different layers of the edition.

#### 4. Documentary and Textual Encoding

19 This way out is not simply a reaction to encoding problems. It is closely related to the idea that every document must be considered a material object on the one hand and a medium of textual transmission on the other (cf. Brüning, Henzel, and Pravida 2012). This distinction has deep roots in the scholarly tradition, but the actual use in editorial theory is due to Hans Zeller's authoritative article, "Record and Interpretation," whose German original dates from 1971. It is known to anglophone scholars in its English translation of 1995 (Zeller 1995).<sup>6</sup> Zeller distinguishes between the manuscript material encountered by the editor—the "record" (*Befund*)—and its "interpretation" (*Deutung*, in a somewhat strict sense of the term; for the sake of clarity, we will speak of "editorial interpretation"). Following Zeller, it is the task of the editor to document manuscripts as objectively as possible and to give an editorial interpretation of them in the form of a well-defined representation of the text and of the genetic process that led to its present form (Zeller 1995, 44). In the German tradition of textual scholarship, observance of this distinction has proven to be the touchstone for editions claiming to attain the highest scholarly standards available.

20 The genetic edition of Goethe's *Faust* aims to fulfill both objectives. The first aim is to make all of the material relating to the work's genesis accessible by means of digital images. To ensure that the record is not only available but also easily accessible, every image is accompanied by a diplomatic rendering which preserves the original lineation and all other relevant inscriptional features of each manuscript and which can be used as a device for close scrutiny of the facsimile or of the manuscript itself. The diplomatic rendering is based on a documentary transcript (example 2) which describes the spatial arrangement of the inscription, scribal hands and writing medium, manipulations of the written letters (for example different forms of deletions), and special marks such as proofreading symbols and print instructions. The documentary transcript does not allow for any textual information on the structural divisions or textual alterations.<sup>7</sup>

21 The second aim is to provide an editorial interpretation of the record from a distinctly textual perspective, with special emphasis on the development of the text brought about by successive alterations and modifications. Various forms of text presentations and of critical apparatuses

are used for the systematic collection, presentation, and analysis of these alterations (example 3). Strictly speaking, the term "textual transcript" is shorthand for "textual and genetic transcript," as the textual transcript serves to satisfy two purposes: to describe the structural subdivisions of the text, scenes, speeches, verse lines, and the like; and to capture the genesis of the text, in particular alterations of the text, bonded variants (in the terminology of Zeller 1995, 49), and also, if identifiable, the phases of the writing process.<sup>8</sup> The textual transcript is kept free from any documentary elements and attributes.

#### 4.1 The Documentary Transcript

22 As example 2 shows, the documentary transcript is designed both to record the sediments of the writing process and to determine the layout of the diplomatic rendering shown to the user.

```
<line type="inter">
  <handShift new="g_bl_lat"/>
  <seg f:left="#Wenn">sie</seg>
  <seg f:left="#erreichen">entschlie&#x17f;sen</seg>
</line>
<line>
  <handShift new="#g_t_lat"/>Wenn<anchor xml:/>
  <handShift new="#g_bl_lat"/><seg rend="inbetween">&#x17f;</seg><handShift new="#g_t_lat"/>ich <mod
    hand="#g_bl_lat" rend="strikethrough">es <anchor xml:/>erreichen</mod>
  könnte </line>
```

Example 2: Documentary transcript of figure 1.

23 The elements used require some further comment because neither the elements themselves nor the encoding approach in general could be conceived of under version 1.3.0 of P5 (TEI Consortium 2008). Even taking into account the new encoding structures recommended in the revisions to chapter 11 of version 2.0.0 of P5 (TEI Consortium 2011a), the encoding above still may seem controversial in certain respects. Considering the type of the markup used, three different kinds of elements and attributes may be distinguished:

1. well-established elements and attributes found in version 1.3.0 of P5 (<handShift/>, <seg>, <anchor/>, @type, @new, @xml:id, @rend, and @hand)
2. elements introduced in version 2.0.0 of P5 (<line> and <mod>)
3. customized elements and attributes—that is, elements and attributes designed for the specific purposes of the edition (in this case <f:left>)

24 The elements and attributes of the first group are commonly used: the values of @new and @hand ("g\_bl\_lat" and "g\_t\_lat") combine the values of the three attributes scribe="g" (Goethe), medium="bl" (lead pencil) or medium="t" (ink), and script="lat" (Latin).

25 The elements of the second group have only recently been incorporated into the TEI Guidelines and deserve a close look. The new element <line> is analogous to the well-established <p> and <l> but is defined in strictly spatial terms (cf. TEI Consortium 2011a, 1083). Our encoding rules demand that a document be transcribed from top to bottom and from left to right according to the spatial distribution of the written characters on the page. The penciled words "sie" and "entschließen" in figure 1 form an interlinear line of their own. In the old version of chapter 11, the relative spatial position of subsequently inserted characters was indicated solely by means of the attribute @place (<add place="above">sie</add> and <add place="above">entschließen</add>), which proved to be insufficient for more exacting diplomatic purposes. Thanks to the newly introduced <line> element, the manuscript lineation, including interlinear lines, can be encoded directly and explicitly. Instead of a structural element like <l>...</l> or <p>...</p> containing the element <add place="above">...</add>, there are two (or more) separate <line>...</line> elements, one for the original line (in figure 1 "Wenn ich es erreichen könnte"), the other(s) for ulterior additions above or below the line, at the left or right margins, or anywhere else. However, the TEI Guidelines still allow the combination of <line> with @place.<sup>9</sup> As a consequence, it becomes possible to inconsistently mix structural and spatial markup and, by implication, to confuse a "document-focused" and

a "text-focused" approach. For example, even though an interlinear line can be considered as part of a preceding or following line, additions in the margin or at the top or bottom of the page should not be treated in that way. Adding <line> to the list of well-established transcriptional elements and attributes may therefore lead to inconsistencies.<sup>10</sup> Another new element is <mod>. As compared with <line>, the use of the <mod> element supplied with a @rend value of "strikethrough" instead of <del> may seem overly meticulous or indirect. But in cases where any suggestions about the nature and coverage of the alteration are to be avoided, the element <mod> is indispensable. (As explained below, the respective scopes of a positively visible cancellation and a textual alteration do not necessarily coincide.)

26 Elements and attributes of the third group (marked with the namespace prefix "f" for *Faust*) have been introduced where the existing ones did not meet the requirements set by our editorial principles. To encode figure 1, the horizontal position of the interlinear line has to be indicated. The spatial relation to the following line has to be expressed straightforwardly, not indirectly (as, for example, by means of a common coordinate system).<sup>11</sup> The spatial distribution of the inscription is not mimicked but represented in an idealized (or, as we prefer to say, structured) way.<sup>12</sup> In example 2, the left margin of "entschließen" is left-justified with respect to the left margin of "erreichen", although figure 1 shows a minimal horizontal displacement. The penciled "sie" was clearly intended by the writer to be placed above the inter-word space between "Wenn" and the original "ich"; mere imitation of the appearance of the source document would end up in locating "sie" above the closing part of "Wenn".<sup>13</sup> To express the obtaining spatial relations, the left margin both of the inter-word space and of the struck-out "erreichen" is marked with <anchor/>. The attribute @f:left indicates that the left margins of <seg>sie</seg> and of <seg>entschließen</seg> are to be left-justified with respect to the <anchor/> element to which the respective value of @f:left points. The diplomatic rendering is thereby fully determined by the encoding.

## 4.2 The Textual Transcript

27 From a textual perspective, however, the very same piece of manuscript inscription looks distinctly different. Here, the characters in ink and in pencil are all to be interpreted as forming one single verse line in which a joint substitution occurs:

```
<l>Wenn <subst>
  <del>ich es erreichen könnte</del>
  <add>sie &#x17f;ich entschlie&#x17f;sen könnte</add>
</subst></l>
```

Example 3: Textual transcript of figure 1 (simplified).

28 The textual transcript looks much simpler than the documentary one. This is not because the textual perspective is generally less complex (sometimes quite the opposite is the case).<sup>14</sup> Nor is it because two attributes have been left out for the sake of brevity.<sup>15</sup> What happens to the passage? A verse line that first read "Wenn ich es erreichen könnte" (if I could achieve [my goal]) now reads "Wenn sie sich entschließen könnte" (if she could bring herself [to do something]). From a textual perspective, as explained above, one personal pronoun, "ich" (I), is substituted by another pronoun, "sie" (she), as the letters s and i, c, h make up the reflexive pronoun "sich" and join the verb "entschließen" to substitute for "es erreichen"; "könnte", i.e., the first person singular conjunctive preterit of *können* (if I could), is substituted by the homograph third person form "könnte" (if she could). From a documentary perspective, however, "könnte" is as unaffected as "Wenn".

29 To put it more generally: The basic elements of the documentary perspective are *written letters*, whereas the basic elements of the textual perspective are *linguistic units*.<sup>16</sup> Depending on the perspective chosen, one and the same phenomenon has to be treated in two different ways. The textual transcript, thus, is neither more nor less adequate than the documentary one. They are categorically distinct: the first records the sediments of the writing process, the second records the genesis of the text.

## 5. Correlating the Documentary and the Textual Transcript

30 The outlined encoding approach shares some similarities with the first method of handling non-hierarchical information as outlined in section 20.1, "Multiple Encodings of the Same Information," of the TEI Guidelines (TEI Consortium 2011a, 604–605). This way of proceeding is, apart from its own merits in terms of theory and sound editorial practice, "the simplest method of disentangling two (or more) conflicting hierarchical views of the same information [...], each time capturing a single view" (ibid.). Each data file can be processed for the purpose of automatic rendering of different types of visualization, as described above.

31 However, such division of the encoding suffers from the disadvantage "that the method requires the maintenance of multiple copies of identical textual content (an invitation to inconsistency) and that there is no explicit indication that the various views, which might be in separate files, are related to each other: it might prove difficult to combine the views or access information from one view while processing the file that contains the encoding of another" (TEI Consortium 2011a, 605). Before we address the problems mentioned here, we would like to emphasize that the disentangling does not result in "multiple copies of identical textual content;" it leads to only a *partial* duplication of the marked-up content.<sup>17</sup> Whereas the documentary and the textual perspective must be considered as two sides of the same coin, i.e., as two different perspectives on one and the same written text, they cannot be regarded as encodings of the very same information. The textual transcript differs from the documentary transcript not only in terms of markup (as is presupposed in chapter 20 of the TEI Guidelines), but also in terms of the marked-up content: "sie entschließen Wenn sich es erreichen könnte" versus "Wenn ich es erreichen könnte sie sich entschließen könnte" (see figs. 4 and 5).

32 Yet there is identical content in these two TEI documents and with it the threat of inconsistency. The resulting task is twofold:

1. The parts of each literal sequence which *should* be identical are to be identified (that is, they have to be distinguished from those parts that are consciously designed to vary).
2. It has to be determined whether these parts *are* identical or not (that is, unintended variation must be detected).

33 As in the passage under discussion, variations mainly occur in the context of alterations.<sup>18</sup> When there are no alterations, both literal sequences should be plainly identical (this is the case for the initial "Wenn"). Where alterations occur, as in the rest of the line, things are more complex. As far as the deleted wording in the textual transcript is concerned ("ich es erreichen könnte", tagged with the <del> element), the marked-up content of this transcript still has a direct counterpart in the documentary one. This requires isolating the string "ich" in the documentary transcript (instead of "sich"), which is possible if the fact is taken into account that the "s" is tagged as inserted later than the original "ich" (see example 2).

34 More serious problems arise when the correspondence between interlinear lines and inserted characters (documentary perspective) and additions (textual perspective) is to be detected automatically. In the textual transcript the strings "sie" and "entschließen" (tagged with the <add> element) follow the string "Wenn ich es erreichen könnte", whereas in the documentary transcript they precede it (see example 4). A collation algorithm which handles transpositions would be able to detect the correspondence.<sup>19</sup>

```
<line>sie entschlie&#x17f;sen</line> <line>Wenn <seg>&#x17f;ich es erreichen könnte</line>
<l>Wenn <subst><del>ich es erreichen könnte</del><add>sie &#x17
```

Example 4: Correspondences between the documentary and the textual transcript.

35 An algorithm that fulfills these requirements is presently being developed by Ronald Haentjens Dekker and Gregor Middell (2011). But even if we assume its availability, the textual transcript captures another linguistic level (words, word groups, etc.) beyond the documentary transcript (graphemes). For example, an insertion of a character in the documentary transcript may correspond to the replacement of a whole word in the textual transcript. In our example, the inserted "f" in the added "sich" in the textual transcript, which is tagged with the <add>



element, lacks a direct counterpart in the documentary transcript: here, the "<seg>f</seg>ich" has been segmented into the strings "f" and "ich". The correspondent of the documentary "ich" in the textual transcript is "<del>ich ...</del>" (see example 3), and if the one-to-one correspondence between the corresponding pairs of strings is preserved, only the "f" in the documentary transcript remains as a potential counterpart of "<add>... f...</add>" in the textual transcript. The same holds true for the added third person form of *können*—"könnte"—the occurrence of which does not leave the slightest trace in the documentary transcript.

36 At this point the automatic correlation reaches its limits. Manual post-processing of collation results is, as always, necessary. For most parts of the encoding, however, the feared inconsistency can be kept effectively under control by automatic collation. And what is more, the same tool which allows for the reduction of unintended variation will also help "to combine the views or access information from one view while processing the file that contains the encoding of another" (TEI Consortium 2011a, 605).

## 6. Conclusion

37 When using a scholarly edition, the user will either read the final text and want to look up all earlier versions of a given line and their manuscript source, or will want to follow the genesis of the text or one of its passages from its germ. To attain this aim, all parts of the edition have to be closely connected, and the available connections have to be intuitively and intelligibly visualized (for more details, see Bohnenkamp et al. 2011). Correlating documentary and textual transcripts, as described in this paper, is only the first step towards the general goal of affording the user good ways to navigate between corresponding parts of the digital archive, the genetic edition, and the established text.

---

### Bibliography

Bohnenkamp, Anne. 1994. "... das Hauptgeschäft nicht außer Augen lassend." *Die Paralipomena zu Goethes 'Faust'*. Frankfurt am Main: Insel.

Bohnenkamp, Anne, Gerrit Brüning, Silke Henke, Katrin Henzel, Fotis Jannidis, Gregor Middell, Dietmar Pravida, and Moritz Wissenbach. 2011. "Perspektiven auf 'Faust.' Zur historisch-kritischen Hybridedition des 'Faust.'" In *Jahrbuch des Freien Deutschen Hochstifts*, 23–67.

Brüning, Gerrit, Katrin Henzel, and Dietmar Pravida. 2012. "On the Dual Nature of Written Texts and Its Implications for the Encoding of Genetic Manuscripts." In *Digital Humanities 2012. Conference Abstracts*, edited by Jan Christoph Meister, 131–134. Hamburg: Hamburg University Press.

Buzzetti, Dino, and Jerome McGann. 2006. "Electronic Textual Editing: Critical Editing in a Digital Horizon." In *Electronic Textual Editing*, edited by Lou Burnard, Katherine O'Brien O'Keefe, and John Unsworth, 51–71. New York: MLA.

Driscoll, M. J. 2006. "Levels of Transcription." In *Electronic Textual Editing*, edited by Lou Burnard, Katherine O'Brien O'Keefe, and John Unsworth, 254–61. New York: MLA.

Gabler, Hans Walter, George Bornstein, and Gillian Borland Pierce, eds. 1995. *Contemporary German Editorial Theory*, 17–58. Ann Arbor: University of Michigan Press.

Goethe, Johann Wolfgang. 1887–88. *Goethes Werke*, edited under the auspices of the Grand Duchess Sophie of Saxe-Weimar-Eisenach. Vols. 14 and 15, *Faust. Eine Tragödie*, edited by Erich Schmidt. Weimar: Hermann Böhlaus.

Haentjens Dekker, Ronald, and Gregor Middell. 2011. "Computer-Supported Collation with CollateX: Managing Textual Variance in an Environment with Varying Requirements." Paper presented at Supporting Digital Humanities, Copenhagen, November 17–18, 2011.

Kraft, Herbert. 1990. *Editionsphilologie*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Pierazzo, Elena. 2009. "Digital Genetic Editions: The Encoding of Time in Manuscript Transcription." In *Text Editing, Print and the Digital World*, edited by Marilyn Deegan and Kathryn Sutherland, 169–86. Farnham, England: Ashgate.

Pierazzo, Elena. 2011. "A rationale of digital documentary editions." In *Literary and Linguistic Computing* 26(4): 463–77. doi:10.1093/lc/fqr033

Renear, Allen, Elli Mylonas, and David Durand. 1996. "Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies." In *Research in Humanities Computing* 4:263–80.

- Saller, Harald. 2003. "HNML – HyperNietzsche Markup Language." In *Jahrbuch für Computerphilologie* 5:185–192. Accessed March 9, 2012. <http://computerphilologie.uni-muenchen.de/jg03/saller.html>.
- Sattler, D. E. 1975. "Vorwort." In Friedrich Hölderlin, *Sämtliche Werke*. Frankfurter Ausgabe. Historisch-kritische Ausgabe, edited by D. E. Sattler, *Einleitung*, 9–19. Frankfurt am Main: Roter Stern.
- Sattler, D. E., and Wolfram Groddeck. 1977. "Schriftarten, Zeichen und Abkürzungen." In Friedrich Hölderlin, *Sämtliche Werke*. Frankfurter Ausgabe. Historisch-kritische Ausgabe, edited by D. E. Sattler, 8:8–9. Frankfurt am Main: Roter Stern.
- Stetter, Christian. 1997. *Schrift und Sprache*. Frankfurt am Main: Suhrkamp.
- TEI Consortium. 2008. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, edited by Lou Burnard and Syd Bauman. Version 1.3.0. Oxford: TEI Consortium. Last updated February 1, 2009. <http://www.tei-c.org/Vault/P5/1.3.0/doc/tei-p5-doc/en/Guidelines.pdf>.
- TEI Consortium. 2011a. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, edited by Lou Burnard and Syd Bauman. Version 2.0.1. Charlottesville: TEI Consortium. Last updated December 22, 2011. <http://www.tei-c.org/Vault/P5/2.0.1/doc/tei-p5-doc/en/Guidelines.pdf>.
- TEI Consortium. 2011b. "P5 Version 2.0 Release Notes." Accessed January 18, 2012. <http://www.tei-c.org/release/doc/tei-p5-doc/readme-2.0.html>.
- Van Hulle, Dirk. 2004. *Textual Awareness: A Genetic Study of Late Manuscripts by Joyce, Proust, and Mann*. Ann Arbor: University of Michigan Press.
- Vanhoutte, Edward. 2002. "Putting Time back in Manuscripts: Textual Study and Text Encoding, with Examples from Modern Manuscripts." Paper presented at the ALLC/ACH, Tübingen, July 25, 2002. Accessed March 9, 2012. <http://www.edwardvanhoutte.org/pub/2002/allc02abstr.htm>.
- Workgroup on Genetic Editions. 2010. "An Encoding Model for Genetic Editions." Accessed January 13, 2012. <http://www.tei-c.org/Activities/Council/Working/tcw19.html>.
- Zeller, Hans. 1995. "Record and Interpretation: Analysis and Documentation as Goal and Method of Editing." In *Contemporary German Editorial Theory*, edited by Hans Walter Gabler, George Bornstein, and Gillian Borland Pierce, 17–58. Ann Arbor: University of Michigan Press.

---

## Notes

- 1 The edition is edited by Anne Bohnenkamp (Freies Deutsches Hochstift, Frankfurt), Silke Henke (Goethe- and Schiller-Archive, Weimar) and Fotis Jannidis (University of Würzburg); see <http://www.faustedition.de>. A detailed account of all the basic parts of the new edition of Goethe's *Faust* and its editorial approach is given in Bohnenkamp et al. (2011).
- 2 For a brief account of the history of research on the genesis of Goethe's *Faust*, see Bohnenkamp et al. (2011, 27–31).
- 3 Recent projects such as *Vincent van Gogh: The Letters* (<http://www.vangoghletters.org/vg/>) and *Walt Whitman Archive* (<http://www.whitmanarchive.org/>) pursue similar purposes.
- 4 At the same time, the "existing chapter on Physical Transcription [i.e. chapter 10] has been exhaustively revised to cater for these revisions which will, it is hoped, go some way to improve the support offered by the TEI scheme for documentary and genetic editing" (TEI Consortium 2011b).
- 5 The term "sediments of the writing process" is adopted from Stetter (1997).
- 6 Throughout this text we often gratefully adopt the solutions found for the translation of German editorial terminology in Gabler, Bornstein, and Pierce (1995).
- 7 This distinction is not to be confused with M. J. Driscoll's "levels of transcription" (Driscoll 2006, 254). In Driscoll's model, the "diplomatic transcription" and the normalized transcription are opposite extremes on a continuous scale; our model presupposes a conceptual duality between the diplomatic and the textual domain.
- 8 D. E. Sattler coined the term "phase" for a temporal segment of the writing process. A segment is defined by operations of extension or alteration of an evolving text. Criteria for the assumption of a phase are material cues such as cancellations or changes of scribal hand (see Sattler 1975, 18–19). At the inception, Sattler intended the concept of phases (*Textphasen*) to hold for the whole of his edition, but he dropped it in subsequent volumes. The

editorial solutions adopted in the introductory volume came in for massive criticism, precisely concerning the editor's application of the very concept of phases. But the concept—if not Sattler's particular way of applying it—definitely deserves reconsideration.

9 See the XML examples corresponding to figures 14, 16, and 17 in section 11.3.4 of the TEI Guidelines (TEI Consortium 2011a, 368, 370).

10 Conceptually similar phenomena, e.g., interlinear and marginal additions, would have to be treated in completely different ways. Alternatively, one would be compelled to constantly sway between the "document-focused" and the "text-focused" approach (TEI Consortium 2011b) by encoding written characters as part of "topographic" lines which they are not part of (`<line>a <add place="bottom">b</add> c</line>`).

11 Cf., however, section 11.4 of the TEI Guidelines (TEI Consortium 2011a, 371–74).

12 This approach is, in a sense, suggested by the principle of preserving spatial relations with structural import ("strukturelle Räumlichkeit") as has been suggested by Herbert Kraft (cf. Kraft 1990, 112–113).

13 The documentation of the record, therefore, involves a good deal of editorial interpretation, a fact of which Zeller is well aware (Zeller 1995, 44). As he demonstrates, the differentiation between descriptive and interpretative information about the record is nevertheless reasonable (ibid.). That is, Zeller's terminological differentiation does not imply that it would be possible to talk about the record without the interference of any kind of interpretation whatever (cf., in contrast, Workgroup on Genetic Editions 2010, section 1.1; Pierazzo 2011, 465).

14 It should be kept in mind that the example given is fairly simple. There is no need for the elements `<undo>`, `<redo>`, `<transpose>`, or the new attribute `@change` introduced in version 2.0.0 of P5. A clarification of their use is urgently needed, but a discussion would exceed the limits of this article.

15 The `<l>` is provided with a reference to the standard line numbering system (in this case line 5142 of the final text). The `<subst>` is classified with regard to the temporal position within the writing process: we always determine whether a revision occurred either immediately after, or still during, the same phase of the writing process ("soon"), or only after the complete text has been written down ("late").

16 In the field of digital humanities the word "text" is usually associated closely with the way texts are digitally represented. "Text" has thus come to be defined as an ordered hierarchy of content objects (see Renear, Mylonas, and Durand 1996) or a strictly linear sequence of characters (see Buzzetti and McGann 2006, 60). If these definitions were adequate to cover the notion of (written) text, our differentiation between a documentary and a textual transcript would not be reasonable, because each transcript might be considered with equal justification as an ordered hierarchy of content objects and as a strictly linear sequence of characters. By maintaining the differentiation we emphasize the genuinely linguistic dimension of written texts that should not, by principle, be collapsed into the dimension of its physical dimension, be it analogue or digital.

17 For the reasons mentioned in the previous footnote we prefer the term "marked-up content" instead of the *Guidelines*' "textual content" (TEI Consortium 2011a, chapter 20.1, 605).

18 Other variations are less critical. For example, foliations are transcribed only in the documentary transcript (tagged with the `<fw>` element), whereas editorial emendations are made only in the textual transcript (tagged with the `<corr>` element). The content of these elements has no counterpart in the other transcript, but the resulting variation can easily be suppressed with the help of a filter engine.

19 In this context, "transposition" refers to one of the four kinds of operations needed to transform the string of the documentary transcript into the string of the textual transcript (or vice versa).

---

### *Cite this article*

Electronic reference

Gerrit Brüning, Katrin Henzel and Dietmar Pravida, « Multiple Encoding in Genetic Editions: The Case of "Faust" », *Journal of the Text Encoding Initiative* [Online], Issue 4 | March 2013, Online since 08 March 2013, connection on 03 June 2013. URL : <http://jtei.revues.org/697> ; DOI : 10.4000/jtei.697

---

### ***Authors***

#### **Gerrit Brüning**

Gerrit Brüning is research assistant at the Freies Deutsches Hochstift in Frankfurt on the Main and at the University of Würzburg. He has published and is working on Weimar Classicism as well as the digital edition of Goethe's *Faust*.

#### **Katrin Henzel**

As research assistant of the Klassik Stiftung Weimar Katrin Henzel is working for the digital edition of Goethe's *Faust*. She studied German Literature, Medieval and Modern History as well as General and Comparative Literature at the University of Leipzig and at Université Sorbonne Nouvelle (Paris III). In 2012, she received a Ph.D. (Dr. phil.) from the University of Leipzig with a dissertation about autograph albums from the second half of the 18th century.

#### **Dietmar Pravida**

Dietmar Pravida is research assistant at the Goethe-Universität Frankfurt am Main, Institute for German literature. He has published a number of studies, critical editions and commentaries on authors of German Romanticism. His current research activities focus on the digital edition of Goethe's *Faust*.

---

### ***Copyright***

TEI Consortium 2013 (Creative Commons Attribution-NoDerivs 3.0 Unported License)

---

### ***Abstract***

The aim of the present paper is to show how, and to what extent, the standards of critical genetic editions as applied to Goethe's *Faust* can be attained within a TEI framework. It proposes and argues for the introduction of two separate transcripts: documentary and textual. Despite the apparent disadvantages of multiple encoding, this approach recommends itself for practical reasons (e.g., avoidance of overlapping hierarchies), and it conveniently reflects the idea that any written document must be considered a material object on the one hand and a medium of textual transmission on the other. In the course of the paper, some aspects and problems of chapter 11 of version 2.0.0 of TEI P5 (the definition and use of the elements <line> and <mod> and related issues) will be discussed.

### ***Index terms***

**Keywords** : genetic encoding, multiple encoding, manuscript transcription, diplomatic transcript, overlapping hierarchies, physical structure vs. textual structure