

# The German Hamlets: An Advanced Text Technological Application

**Benjamin Birkenhake**

University of Bielefeld, Germany

**Andreas Witt**

andreas.witt@uni-tuebingen.de

University of Tübingen, Germany

The highly complex editorial history of Shakespeare's Hamlet and its many translations to German forms a interesting corpus to show some of the major advantages current texttechnologies. In the following we want to present the corpus of four English editions an several German translations of the play we have gathered together and annotated and crosslinked in different ways.

Although Shakespeare's Hamlet is obviously not a unique hypertext, it is an interesting object to test advanced hypertext- and text technologies. There is no original edition of Hamlet, which was authorized by Shakespeare during his lifetime. We only have different print editions, which all have a different status concerning their quality, overall length, content and story-line. The most important among these are the so called first folio, the first quattro and the second quattro edition of Hamlet. During the centuries editors tried to combine these early editions to the best edition possible. The famous Arden-editions as well as the in the internet widespread Moby-edition are such compositions.

A comparable but a bit more complex situation exists within in the field of german translations of the play. The earliest translation is by Christoph Martin Wieland from about 1766. After this at least 18 translations have been published which are accompanied by countless translations of theatre directors, which are mostly not documented. The corpus contains 8 digitalized Translations. 2 further translation are already scanned but not yet digitalized, because they are printed in fraktur - a old german typeface - which can not be recognized by common OCR-programs yet. The remaining 10 Translations are available in print, but not yet digitalized, too. Of the 8 digitalized translations we chose 4 for further text technological use.

What makes the corpus so interesting is the fact, that almost every translator used several of the early english editions as a basis for a new translation. This leads to a situation in which almost every german or english edition of Shakespeare's Hamlet is a composition of several sources. The relation the editions have with their sources and with each other form a wide network, which could be presented in a hypertext system.

Another interesting aspect of Shakespeare's Hamlet is the outstanding position the play has within the western culture for centuries. Hamlet is the single most researched piece of literature, has been analyzed from various perspectives and is a part of western common education. This leads to the request, that a digital environment should represent the variety of perspectives on the play. This lead us to a corpus of Hamlet editions in which each text may exist in multiple forms.

Basis for the XML-annotations are text files, which are transformed to XML using regular expressions. The basic XML-format is TEI 4 drama base tag set. TEI 4 is a major open source concept of the Text Encoding Initiative. The drama base tag set offers almost all tags needed for a general, formal annotation of a play. In order to provide an easy to annotate mechanism we added some attributes to represent the translation- or origin-relation between lines, paragraphs or speeches within the editions on the one hand and the sources on the other hand.

The TEI-annotated documents are used for further annotations an presentation. The TEI-documents were automatically enriched with further markup, using an open source auto-tagger. This auto-tagger annotates single words, including the part of speech and the principle form. The TEI-documents are also the basis for the XHTML-presentation. As the TEI structure contains all information necessary for a graphical presentation, these documents are transformed to XHTML, which is used to present the corpus. This transformation is made with several XSLT-Stylesheets. In the same way XSLFO is used to generate PDF-versions of each edition.

edition	Txt	TEI	XHTML	STTS	Narration
1st Folio	x	x	x		
1st Quattro	x	x	x		
2nd Quattro	x	x	x		
Moby	x	x	x		
Wieland	x	x	x		x
Schlegel	x	x	x	x	x
Fontane	x	x	x	x	x
Hauptmann	x	x	x	x	

Table 1: The different layers of the mayor editions within the corpus

In many cases translators have re-arranged the flow of stanzas or the course of action. Therefore it is useful to provide an alternative linking mechanism, which does not only focus on the language and the formal structure, but also on the plot. To provide this reference the narrative information is annotated in another layer. This allows to find the same event in different translations of the play. The narrative annotation layer basically consists of events, which can be seen as the smallest elements of the plot.

Obviously, events may start within one line and end several lines or even speeches later. Since the narrative structure is overlapping with the TEI, both are stored in separate

annotations. Scenes can provide a meaningful unit for basic parts of the plot. Thus the formal and the narrative annotation are semantically aligned - in addition to their reference on identical textual data. This relation can be exploited by creating links between the concept of a scene and the concept of specific actions. The respective linking mechanism is located on a meta level: it operates on the schemas themselves and not on their instances. The references are generated mechanically on the meta level, linking different perspectives together. Readers can explore the relations between events and scenes. The procedure could also be used to create a recommendation system as e.g. proposed by Macedo et al. (2003): the annotation integrates the knowledge of experts on narrative structures in the play Hamlet and provides this information to the reader. This leads to a multi rooted tree, each tree represents one level of information, i.e. textual structure and linguistic, philological or narrative information. This allows for creating a network of multiple perspectives on one text being linked to one another. As a result, hypertext is no longer based on links between nodes, but offers a reference mechanism between perspectives.

other. In the first case, the annotation of events and actions provides a way of comparing different editions esp. translations.

4. Using SVG - an XML-based format for graphics - the narrative structure of each translation could be visualized, ignoring the textual basis. This gives an »overview« of plot of the current edition.

5. The introduced concept of cross annotation linking allows us to offer the user automatically generated links from one annotation to another.

With this set of different linking-concepts we offer users a new freedom to explore the corpus in a way that fits to their needs. We ensured, that every layer of information offers a way to access information of another layer in a different perspective. We assume that this method can be transferred to any kind of multiple annotation.

## References

Macedo, A. A., Truong, K.N. and Camacho-Guerrero, J. A. (2003). Automatically Sharing Web Experiences through a Hyperdocument Recommender System. In: Proceedings of the 14th conference on Hypertext and Hypermedia (Hypertext03), (Nottingham, UK, August, 26-30, 2003). Online: <http://www.ht03.org/papers/pdfs/6.pdf> [Available, Last checked: 6/16/2005]

```
<event>
  <event unit="u0024" sequel="u0030" >HORATIO: Das kann ich, wenigstens das Gerede geht so.</event>
  <event unit="u0013" sequel="u0015" >Unser letzter König, dessen Bild soeben vor uns erschien,</event>
  <event unit="u0025" sequel="u0026" >ward wie ihr wißt -vom Norweg Fortinbras - durch welt-
  etfernen Stolz dazu gespönt - zum Kampf gefordert, </event><event unit="u0026" sequel="u0027" >in
  welchem unser tapfrer Hamlet (nach unsrem Wissen
  schätzte ihn die Welt von dieser Seite) den Fortinbras
  schlug.</event><event unit="u0025" sequel="u0026" > welcher - nach einem festgestellten Verträge - [...] </event>
</event>

<sp nr="46" ftr="q1r" q2r="" moby="46" who="">
  <speaker>HORATIO</speaker>
  <ln="91" ftr="" q1r="" q2r="" moby=""> Das kann ich, wenigstens das Gerede geht so.</ln>
  <ln="92" ftr="" q1r="" q2r="" moby=""> Unser letzter König, dessen Bild soeben vor uns erschien,</ln>
  <ln="93" ftr="" q1r="" q2r="" moby=""> ward wie ihr wißt -vom Norweg Fortinbras - durch welt-</ln>
  <ln="94" ftr="" q1r="" q2r="" moby=""> etfernen Stolz dazu gespönt - zum Kampf gefordert, in</ln>
  <ln="95" ftr="" q1r="" q2r="" moby=""> welchem unser tapfrer Hamlet (nach unsrem Wissen</ln>
  <ln="96" ftr="" q1r="" q2r="" moby=""> schätzte ihn die Welt von dieser Seite) den Fortinbras</ln>
  <ln="97" ftr="" q1r="" q2r="" moby=""> schlug, welcher - nach einem festgestellten Verträge - [...]
  [...]
</sp>
```

Figure 1: A multi rooted tree above a single textual data

As a first result of these multiple annotations, we got a corpus that is based on XML-technology and available via the web. As a second result we developed methods to cope with multiple annotated documents, which is a task, that has to be performed more often with the growing popularity of XML-technologies. Especially the integration of the narration annotation layer has to be seen as a example for further parallel annotations. In detail these methods described above lead to an environment, which offers different types of user different perspectives on a single, textual object or a corpus. Some of these benefits will be presented in the following:

1. The common TEI-annotation allows a structural linking-mechanism between the editions. This allows a user to jump from the first scene in the second act of one edition to the same scene in another edition.
2. Alternatively this annotation can be used to present the user a part of the play in on or more editions of his choice for direct comparison.
3. The narration annotation layer allows several ways to explore a single text or compare some texts with each