Reconstructing the textual evolution of a medieval manuscript

Malte Rehbein
Moore Institute, National University of Ireland, Galway, Ireland

Abstract

This article presents the results of the work on *kundige bok*, one of Göttingen's town records, containing late medieval town law. Due to the fact that this law was frequently subject to change, the text itself was revised over and over again, giving evidence for its frequent use and its dynamic nature. What has come to us, is, thus, a multi-layered text in which all layers represent a different (e.g. chronological) stage of the town law. Consequently they have to be regarded, processed and represented equally. A dynamic text like this requires a dynamic representation. The article shows how an electronic scholarly edition of a multi-layered text can be created and used, first, to reconstruct the genesis of the text; second, to make this evolution understandable, processable and visible; and third, with the text as a witness to display the development of urban law and urban life in the Late Middle Ages.

This article:

- outlines the challenge of editing a multi-layered medieval manuscript;
- discusses why this leads to a new understanding of a critical edition of such a text; and
- introduces the techniques used to create the electronic edition of *kundige bok*, in particular highlighting the linkage between the two dimensions of 'text'and 'time' based on the TEI P5 scheme.

Correspondence:
Malte Rehbein, Moore Institute, National University of Ireland, Galway, University Road, Galway, Ireland.
E-mail:
malte.rehbein@nuigalway.ie

1 Background

For this study,[1] we must return to Germany's late fifteenth century to find, somewhere well-positioned along major medieval trade routes, the city of Göttingen.[2] With its 4,000–6,000 inhabitants, Göttingen had to be counted as one of Germany's bigger medium-sized towns, member of the Hanseatic League, and economically well-established due to supra-regional importance, trade, and linen production.[3]

In the previous 200 years, the town had emancipated itself from the rule of the Brunswick Dukes and gained a certain autonomy by acquiring rights such as jurisdiction and the right to mint and issue coins. In the fifteenth century this process of emancipation had come to an end. The town was now self-governing and the Duke's rule had been replaced by a city council. A major feature for this autonomy was town law. The text in question, the *kundige bok* (Book of Announcements), was part of this new law that the town had established for itself. It contains regulations about everyday life in town[4] which the city council had found necessary to announce regularly to the public. The by-laws themselves and the compulsory assembly of the citizens were both called *burspraken*.[5]

Literary and Linguistic Computing, Vol. 24, No. 3, 2009.

doi:10.1093/llc/fqp020 Advance Access Published on 2 June 2009

The assembly was originally part of the *Schwurgemeinschaft* (confederation) of the city's population, but by the fifteenth century it had already been used by the council for a lengthy period as an instrument to demonstrate its authority and to strengthen its rule. It took place in a quite common manner: on a specific date every year the new council was elected upon which it discussed and decided on the regulations for the upcoming year. The population, called by the ringing of the bells, assembled in the town hall square and was obliged to listen to the announcement of the new laws. The council took an active role while the citizens, on the other hand, remained passive. Thus, the *burspraken* indicate clearly who was in power: it was no longer a confederation of equal citizens but a council that symbolized and strengthened its power by means of public address using the *burspraken*.

2 The Text

Kundige bok is a compilation of various texts with the *burspraken*[6] being its main content and used for this study. The book, which was written on paper, is still in quite good condition, written with care regarding the content though not the design: it was a script intended for regular use. It consists of approximately 330 pages written between 1420 and 1500 by various scribes. Some texts were in Latin but the vast amount of text is written in *Mittelniederdeutsch* (middle Low German) which was spoken in North Germany until the sixteenth century.[7]

What is so characteristic about the text that it not only *can* be edited digitally but actually *requires* a presentation as a digital edition? It is the fact that the medieval town law was by its very nature anything but fixed and unchangeable. Over the years, the city council reacted constantly to economic, social or political changes by adapting the by-laws. And, consequently, so did the council's scribes: they frequently modified the written statutes by either altering the existing text or adding completely new versions which could happen as frequently as every year.

The following example (Fig. 1) contains statutes about the brewing of beer.[8] This was indeed an issue of everyday life in medieval Germany since beer was part of the staple diet and was not only produced by commercial breweries for profit but also by the citizens themselves. The various changes to these statutes give evidence of the importance of regulating beer brewing for the city's administration.

The sample page shows these statutes as they were issued and announced between 1484 and 1495: the oldest text appears in the top and centre of the page while more recent alterations can be found at the right margin, below and interlinear.

In the example given (see Fig. 2 for textual evolution), the scribe wrote a draft version of the text. Together with some alterations, making the use of the public mills compulsory, it was initially announced in October 1484. One month later, the council found it necessary to make it easier for new citizens to acquire the right to brew beer, and the text was changed again. Three years later, this regulation was taken back again and replaced by something new. In a final stage, in 1495, the coordination of the brewing days was modified.

Thus, every change made by the medieval scribes in *kundige bok* has to be regarded as a new layer of the text since it had produced a revision of the law by replacing its predecessor. When the law was changed, the previous version of the text becomes obsolete.[9] Each text layer represents a certain stage of the statutes: text became law and law turned into text. We have to regard this page of text not as *one text* but as at least five textual layers of equal importance.

It is, however, important for historical studies or research in the history of law to see not only the version of the text at a given point in time but also to follow its development and its evolution to analyse the changes in its economic, political, and social context. The frequent changes of the text give evidence for a consistency of the town law with its evolutionary modifications. This was, however, not that clear. On the contrary, research so far has mostly regarded the Göttingen by-laws as singularities (being announced only occasionally rather than annually) and has ignored the process in which the city council was active in adopting the regulations to

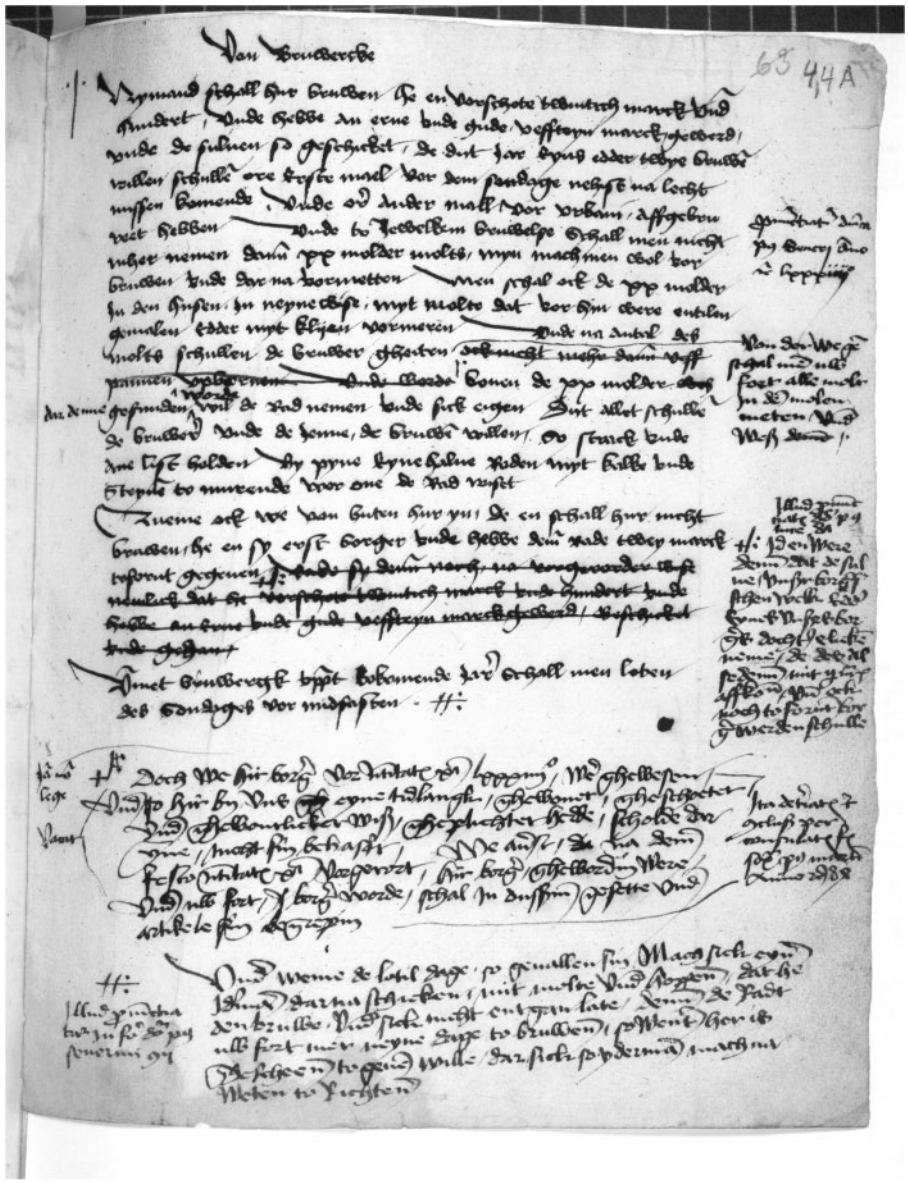

Fig. 1 Sample Page of *kundige bok* (folio KA04_4r)

meet new needs. This is due to the fact that thus far only a conventional edition of parts of the source text exists (Ropp, 1907).

Editing a text like *kundige bok*, in a conventional manner, creating a printed edition of it, meant deciding that one layer of the text was regarded as the most important. This is, however, impossible.[10] It would consequently devalue all the other layers, banning information about them into the apparatus. The user of such an edition will likely follow the editor's decision and might easily neglect the fact that all text layers have to be regarded equally—that they are just as important as any other, only at a different point in history. The static

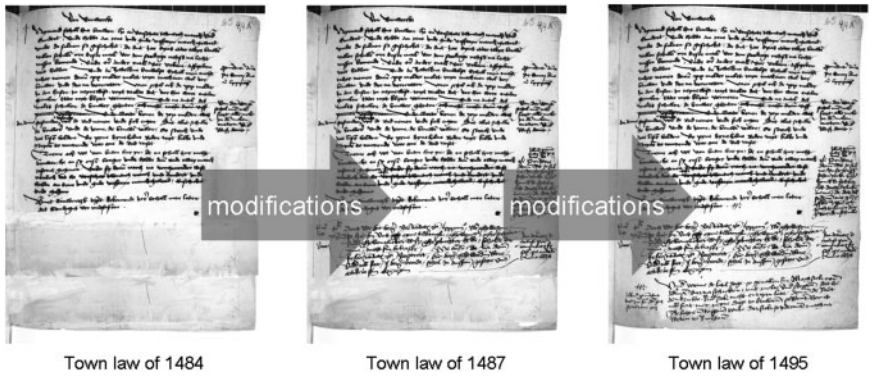

Fig. 2 Textual evolution in *kundige bok*

Fig. 3 Single text layer view and comparison view of the digital edition

format of a printed edition forces the editor to make presumptions about its future readers and thus restricts its use leading to a misunderstanding of the text and to common misinterpretation of town law.[11]

3 The Solution

What interests me having modern technology at hand is: how can you accurately visualize the evolution of the text and consequently of Göttingen's town law? In other words: how can you create a scholarly edition of such a multi-layered text and make it easily accessible to users?

The main tasks for this work are, therefore to:

- find a way to represent the text so that each of its layers is treated equally;
- let the users decide which version of the text is of importance for their own research without any presumptions by the editor that might cause a biased interpretation.

Kundige bok is a dynamic text and it requires a dynamic representation. The decision was, therefore, in favour of not only choosing the computer and the internet as the output medium for the edition, but to create an electronic edition that allows the users to interact with the edition by means of the user-interface and to adapt the output to their research interests.

Within the digital edition of *kundige bok*, the town statutes are reconstructed and easily retrievable for each individual year. It has three main features:

- to give the user an overview of the evolution of the text along the two dimensions of time and law;
- to give the user an access to the text in which the statutes are reconstructed and easily retrievable for each individual year ('single text layer view', see Fig. 3, left);
- to allow any variants of the law to be compared and put into their economic, social, and political context by a simple mouse click ('comparison view', see Fig. 3, right).

Obviously, the dynamic digital edition of this text does not simply imitate a printed edition in a different medium. It is a dynamic approach in which the user actively constructs the edition by his interaction allowing him to gain added value from it.

4 Encoding Multi-layers

4.1 General encoding strategies

Without placing any restrictions on future extensions e.g. linguistic studies, the scholarly edition of *kundige bok* is intended to be used by historians and for the study of legal history. The markup strategy in general is therefore based on having a user in mind that is interested in easily reading the text, especially a reconstruction of city law at a particular point in time while also providing access to its development.

The encoding of *kundige bok* was achieved by following the TEI P5 guidelines and using the P5 scheme without any modifications.[12] The first question to clarify was the leading structure or hierarchy of the XML data. Let us remember: *kundige bok* is characterized not only by its various textual layers but also by the fact that it was regularly used in the Middle Ages and pages had been re-ordered so that the physical structure of the pages as they exist today does not necessarily represent the way the book was used nor does it help to create a readable version of the town law at a certain stage.

Among others, there were three possibilities to structuring the XML data:

- the order of the pages of the book (physical);
- by its textual development (chronological);
- by the topics dealt with (logical).

Since XML in general allows only one hierarchy, the logical structure was chosen as predominant, requiring other mechanisms of encoding its physical structure and chronological development.[13] Thus, the main structure of the document is:

part_of_the_statutes > topic > paragraph

4.2 Encoding challenges

In the following section, the encoding and visualization of the textual evolution will be discussed. There are two major challenges.

First, since chronology is not the leading hierarchy of the document, the simple use of <add> and <del> tags (in conjunction with <subst>) would not work because it overlaps in many cases with the logical structure (mainly because of changes of the paragraphs). Instead, as suggested in the TEI Guidelines, the milestone elements <addSpan> and <delSpan> were used, together with <anchor> elements indicating the end of the span.[14]

However, encoding five or more revisions of the same passage of the text, leads to problems or at least confusion nesting the elements correctly. Not for the computer—the document itself is valid against the P5 schema and very well processable. It is, however, a challenge for the encoder who has to create, understand, and maintain such markup.

Second, in the editorial process, the constitution of the text layers is often not obvious. Questions as to whether one alteration such as an addition, a deletion, a substitution, or a comment is older or more recent than another, whether any two such changes commonly form a text layer or have to be regarded separately, and whether one has to be regarded as an existing text layer at all, need information that is often not at hand while transcribing the script but may become clear at a later stage of the editorial process.

The chronological order of two or more textual changes is indeed relevant in order to understand

the text and without it may lead to misinterpretation. Consider the following[15]:

> *We ock vorschote ~~100~~[150] marck, de darf ~~3~~[2] warve bruwen.*[16]

Without additional information, the following four readings are possible:

R1: *We ock vorschote 100 marck, de darf 3 warve bruwen.*
R2: *We ock vorschote 100 marck, de darf 2 warve bruwen.*
R3: *We ock vorschote 150 marck, de darf 3 warve bruwen.*
R4: *We ock vorschote 150 marck, de darf 2 warve bruwen.*

With more information (mainly palaeographical), three facts become obvious: first, that R1 is the oldest version of the text, second that R4 is its most recent and third that although R2 or R3 may or may not have existed as text layers, both of them could not have existed as a revisionary sequence from R1 to R4.[17] A less formalized way to express this might be: was there any point in time when people were allowed to brew beer twice a year if they were taxed at more than 100 marks? Or was there any year when people were allowed to brew three times if they were taxed at more than a 150 marks? Or did only two versions of the law really exist? Regarding the textual evolution, this allows the following possible 'paths':

p1: R1 > R4
p2: R1 > R2 > R4
p3: R1 > R3 > R4

Again, without more information, all three paths are possible and reasonable. But each of them would lead to a different understanding about the development of the statutes. The main editorial challenge is that, as mentioned previously, it may become clear quite late in the editorial process which path was the actual textual development—if it becomes clear at all. In many cases, there is neither enough palaeographical nor contextual information to make a reliable statement about it.[18] This leads to the necessity of, first, leaving it up to the user to rearrange textual layers wherever the editor is not able to make a reliable statement (thus allowing the user to make up his own mind) and, second, allowing future changes whenever new knowledge is available that permits a more precise definition of the textual development. For both purposes, it is quite important to have the digitized facsimile at hand which is provided at page level within this digital edition.

Overall, a lot of flexibility is needed allowing the editor to make changes in both the definition and order of the text layers, allowing the user to come to his own conclusions where the editor cannot be sure, and ensuring the edition remains open to future changes.

4.3 Splitting up the data

The general idea to solving both problems, making the encoding process as easy as possible and allowing the required flexibility in the definition of the textual layers is to (temporarily) separate the text[19] from the knowledge or data describing the textual development.[20] This approach splits the data into two parts. The first part (the markup) can be completed consecutively, thus creating a 'stable' encoding[21] of any occurrence of text or text-related features within *kundige bok*. For that, the text has to be divided according to each textual change into different units, e.g. one separate unit for every addition.[22] Each unit is encoded by means of the TEI guidelines (Representation of Primary Sources)[23] but has currently no linkage to other units.

In the example above, the editor would build up the following three units:

u1: We ock vorschote 100 marck, de darf 3 warve bruwen.
u2: 150.
u3: 2.

Each of them is at first considered and encoded separately and each of them (except u1) does not really make sense on its own. The linkage in question is realized by the second part of the data which is built up by a lightweight database storing the knowledge about the textual development.[24] This requires the markup to contain anchor tags for identifying the positions where substitutions, additions, deletions, or comments are made.[25]

Consider the first part of the example above. It would be encoded as

We ock vorschote <anchor xml:id=”a1”/>
100<anchor xml:id=”a2”/> marck

while the database would have the following entries:

- change c1: unit u2 is a replacement for unit u1, starts at #a1, ends at #a2;
- R3 and R4 are text layers;
- c1 is member of text layers (or readings) R3, R4.

Obviously, the last two entries in the database are the ones that have to be adaptable until the very end of the editorial process and beyond. They describe the development of text over the years and as mentioned above need to be flexible and easily changeable. Re-ordering the textual layers now simply means changing an entry in the database and not rearranging the whole markup of this section. The database fulfils the requirement of being understandable and maintainable by the human editor[26] without losing its machine-processable attributes.

Since it is processable, it can easily and automatically be transferred into a XML document, valid against TEI P5 and containing all available data. This is done by a script combining both sources of information: the markup itself and the database entries. Changes made during the editorial process now mean altering an entry in the database and simply running the script again. For the example discussed, this script would produce an XML document such as (simplified):

We ock vorschote
<delSpan spanTo=”#a1” seq=”R3,R4”/>
100<anchor xml:id=”a1”/>
<addSpan spanTo=”#a2” seq=”R1”/>
150<anchor xml:id=”a2”/>>
marck

The script produces a document which contains all of the information that was originally held in the database including information about the text layers which are now transferred into the seq attribute.[27] Since it creates a valid TEI file out of the ‘raw data’, it can then be transformed via XSLT and CSS[28] into the desired output. This transformation process needs two inputs: the text layer that is going to be visualized (or two or more text layers in comparison mode) and the interpretative path that is to be used.

Visualizing a certain text layer means applying a transformation rule according to the given path and the text layer in question. Consider an example where the two paths

p’1: R1 > R2 > R3 > R4
p’2: R1 > R3 > R2 > R4

were possible because no information about the real textual development is available. Applying these two paths to construct text layer R2 will necessarily result in different texts. Assuming p’2 as being the development of text and law, would incorporate alterations made in R3 while using p’1 as the path would not. *Kundige bok* shows many examples where the editor cannot provide a reliable path. Thus, the edition’s user interface provides a feature to allow the users to chose by themselves and with the means introduced above it is possible to easily change the definition of the text layers whenever new information has become available.

5 Conclusion

The study of *kundige bok* is an example of a text that has to be constituted not only by its sequence of characters but also by information that it is often not available in the text itself. What is more, this information has to be regarded as dynamic, thus requiring not only a dynamic representation in terms of a digital edition but also a flexible way of encoding and an understanding of an edition where the user may take part actively and which is prepared for future changes.

References

Buzzetti, D. and McGann, J. (2006). Critical editing in a digital horizon. In Burnard, L., O’Brien O’Keeffe, K., and Unsworth, J. (eds), *Electronic Textual Editing*. New York: The Modern Language Association of America, pp. 51–71.

Denecke, D. and Kühn, H. (eds) (1987). *Göttingen: Geschichte einer Universitätsstadt: Von den Anfängen*

bis zum Ende des Dreißigjährigen Krieges. Göttingen: Vandenhoeck & Ruprecht.

Grésillon, A. (2004). *Eléments de critique génétique. Lire les manuscrits modernes*. Paris: PUF.

Huitfeld, K. (2004). Editorial principles of Wittgenstein's Nachlass—the Bergen electronic edition. In Buzzetti, D., Pancaldi, G., and Short, H. (eds), *Augmenting Comprehension: Digital Tools and the History of Ideas*. London: OHC.

Rehbein, M. (2000). *Das olde kundige bok in spätmittelalterlichem Gebrauch*. Magisterarbeit. Göttingen: Universtität Göttingen.

Rehbein, M. (2008). *Göttinger Burspraken im 15. Jahrhundert. Entstehung – Entwicklung – Edition*. Göttingen: Universität Göttingen. PhD thesis (unpublished).

Ropp, G. (1907). *Göttinger Statuten. Akten zur Geschichte der Verwaltung und des Gildewesens der Stadt Göttingen bis zum Ausgang des Mittelalters*. Hannover und Leipzig.

Notes

1 This article reflects the author's work at the University of Göttingen as well as being a member of the TEXTE group. TEXTE is an advanced research programme at NUI Galway, funded through a European Union Marie Curie Transfer of Knowledge Host Fellowship with the aim of exploring the impact of new technologies on textual editing. TEXTE is directed by Prof. Sean Ryder.

2 Approximately 100 km south of Hannover.

3 See Denecke *et al*. (1987) for a number of essays on medieval Göttingen.

4 While we are using the medieval beer production as a case study in this article, the by-laws govern many more issues of everyday life: e.g. taxes, clothing and cloths, coins and currency, rubbish on the streets, fire protection, defence of the city and of the city walls, leisure and gambling.

5 Derived from the low-middle German words *bûr* (inhabitant, citizen) and *sprake* (speech). Ref. *Lexikon des Mittelalters*, 2, 1110–1.

6 In the following, *burspraken* is used for the by-laws, not for the assembly of the town population.

7 Lexikon des Mittelalters, 3, 767–77.

8 Stadtarchiv Göttingen, AB Ms 2,2, folio KA04_4r.

9 There is evidence in the text, however, that previous versions were used quite often for reference purposes.

10 This question is similar to the general discussion about the *critique génétique*. See (Grésillon, 1994).

11 There are many examples of misunderstanding *kundige bok* (Stadtarchiv Göttingen, AB Ms 2,2) and its predecessor *olde kundige bok* (Stadtarchiv Göttingen, AB Ms 2,1 I) which are mostly based on the first (and incomplete) edition of the texts (Ropp, 1907). Providing these examples are not part of this article but they are discussed extensively in (Rehbein, 2000) and (Rehbein, 2008).

12 This was not quite clear at the beginning of the work. But designing the mechanism described below, it turned out that no additions to the tagset were needed.

13 The chronological aspect of *kundige bok* is crucial for this work. For representing the physical structure, the empty elements <lb/> for line breaks and <pb n =''folio_id''/> were simply used, thus requiring some more thinking about the future use of the data (in terms of processing page level related information and transformation into visible output).

14 See Chapter 11.3.4 (Additions and Deletions) of the P5 guidelines: http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html.

15 For a similar example see the discussion on Wittgenstein's manuscripts by (Huitfeld, 2004). See also: http://www.tei-c.org/About/Archive_new/ETE/Preview/huitfeldt.xml.

16 Part of the brewery statutes, regulating the number of times one is allowed to brew according to his tax paying. Roughly translated: one, who pays 100 (150) mark as taxes, is allowed to brew 3 (2) times (a year).

17 'None of them' would mean that the change went directly from R1 to R4, thus all alterations were made at the same time. All four readings could not have existed because there is no path R2 > R3 (from R2 to R3) or R3 > R2 (from R3 to R2).

18 In a number of cases the scribes put a date to the textual changes allowing a precise statement about the textual development. Sometimes palaeographical studies or contextual information helps to narrow the date in question down.

19 In the meaning of sequences of graphemes.

20 For a recent discussion see (Buzzetti and McGann, 2006).

21 Of course, the transcription might change, a different editor might come up with a different reading of a certain word or passage. What I mean here by 'stable' is the fact that the transcription is *relatively* stable and certain (uncertainty will be encoded as such) in comparison to the identification of the textual layers.

22 This forms the highest level on the general hierarchy of the XML data. It does not interfere with part_of_the_statutes > topic > paragraph structure discussed above.
23 Ref. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html.
24 Technically speaking, both sets of data are stored in a database management system. However, only the second set is really database-like for it describes relations and dependencies among the entries.
25 Analogue to the general annotation mechanisms in TEI P5.
26 In the ongoing work, there are 1500+ entries in the database currently.
27 As mentioned above, the milestone elements <delSpan> and <addSpan> are used, because of the leading structure. Changes in the law contain alterations in the arrangement of paragraphs, thus using del and add (nested in a subst) would overlap with it.
28 Extensible Stylesheet Language Transformation and Cascading Style Sheets.