# SusTEInability of linguistic resources through feature structures

**Andreas Witt**
Institut für Deutsche Sprache, Mannheim, Germany

**Georg Rehm**
vionto GmbH, Berlin, Germany

**Erhard Hinrichs**
Tübingen University, General and Computational Linguistics, Germany

**Timm Lehmberg**
Hamburg University, SFB Multilingualism, Germany

**Jens Stegmann**
Bielefeld University, Faculty of Linguistics and Literary Studies, Germany

## Abstract

This article shows that the TEI tag set for feature structures can be adopted to represent a heterogeneous set of linguistic corpora. The majority of corpora is annotated using markup languages that are based on the Annotation Graph framework, the upcoming Linguistic Annotation Format ISO standard, or according to tag sets defined by or based upon the TEI guidelines. A unified representation comprises the separation of conceptually different annotation layers contained in the original corpus data (e.g. syntax, phonology, and semantics) into multiple XML files. These annotation layers are linked to each other implicitly by the identical textual content of all files. A suitable data structure for the representation of these annotations is a multi-rooted tree that again can be represented by the TEI and ISO tag set for feature structures. The mapping process and representational issues are discussed as well as the advantages and drawbacks associated with the use of the TEI tag set for feature structures as a storage and exchange format for linguistically annotated data.

**Correspondence:**
Andreas Witt, Institut für Deutsche Sprache, R 5, 6-13, D-68161 Mannheim, Germany.
**E-mail:**
witt@ids-mannheim.de

## 1 Introduction

This article presents a representation format for the exchange of documents that contain complex markup. It is based on the TEI tag set for encoding feature structures and shows that this TEI tag set not only qualifies as a well-suited meta-format for annotated linguistic corpora, but that it can also serve as

doi:10.1093/llc/fqp024 Advance Access published on 11 June 2009

a method to use XML for the annotation of the otherwise unannotatable.

XML's most fundamental data structure is a tree. While trees have several advantages for software developers as well as for users who mark up textual data, they are able to express nested annotation structures only. The annotations of a document may constitute one or several logical layers, as long as the bracketings within a single layer or across layers never across one another. Linguistically annotated corpora, however, do not necessarily satisfy this constraint. They may contain crossing edges and, thus, require a data structure that is more complex than a simple tree.

Several solutions for this problem have been proposed (see, e.g. DeRose, 2004; Carletta *et al.* 2007). One is to factor such complex and possibly multilayered annotations into a multi-rooted tree, i.e. into several trees spanning over the same leaves. Multi-rooted trees constitute a data structure that is more general than a single tree, but not as unrestricted as an Annotation Graph (Bird and Liberman, 2001). This article shows how multi-rooted trees can be represented in an integrated way, by using the TEI tag set for the annotation of feature structures.

This article is structured as follows: Section 2 presents the underlying technological and methodological framework, i.e. an architecture with the aim of fostering the sustainability of linguistic resources, and describes the task of representing linguistically analysed corpora. Section 3 illustrates the use of the TEI tag set for the representation of feature structures as a storage- and interchange format for multilayer annotations. In Section 4 two alternative approaches on representing multi-layer annotation, XCONCUR and the NITE project format, are briefly described and compared to the feature structure based representation. Section 5 concludes the article with a critical discussion of the practical usability of this approach.

## 2 The GENAU Approach

The work presented in this article is part of a research effort on the sustainability and preservation of language data. A generic framework for assuring the long-term accessibility of heterogeneous linguistic resources was developed within the project Sustainability of Linguistic Data (see, e.g. Wörner *et al.*, 2006; Rehm and Witt, 2007; Witt *et al.*, 2007; Rehm *et al.*, 2008a,b; Rehm *et al.*, 2009). An important aspect of the overall architecture of this project is a specific approach to handling and processing several corpus representation formats, the Generalised Architecture for Sustainability of Linguistic Data (GENAU, see Fig. 1). It includes a mechanism for the representation of complex linguistic corpora and a component for the mapping of linguistic tag sets into an ontology. This article only deals with the representation of corpora, visualized on the right hand side of Fig. 1.

Since linguists investigate corpora from different theoretical points of view, linguistic corpora typically are annotated on multiple levels of description, such as, for example, morphology, syntax, and semantics. To represent these annotations uniformly, the data is XML-encoded concurrently. As a result of this encoding strategy, a separate document instance exists for each annotation level. This approach can be characterized as redundant encoding in multiple forms (Sperberg-McQueen and Burnard, 1994).

However, the redundant encoding of different kinds of information does not account for the fact that there might be interrelations between the different annotation layers. This disadvantage can be avoided if the primary data, i.e. the textual content, is identical across the respective document instances (Witt, 2004). This guarantees that the text functions as an implicit link for the separately realized annotation layers. It should be noted that an approach along such lines is somewhat controversial among members of the markup community. This is mainly due to criticism connected to issues such as data consistency, layer comparison, perceived redundancy, and the availability of seemingly more attractive integrative formats.

From the point of view of sustainability, the multiple encoding approach does have two overwhelming advantages: since the markup/text ratio is relatively low, the XML-encoded files can be used with off-the-shelf XML-software and, furthermore, they are human-readable. Secondly, since linguists

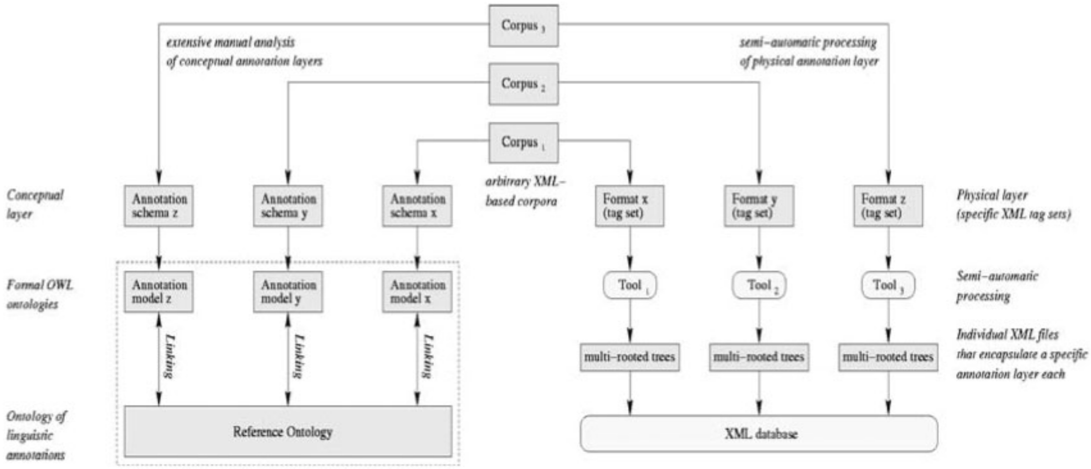


**Fig. 1** The two main phases of the GENAU approach

are often interested in only one (or a small number) of the heterogeneous annotation levels, they can directly access those documents which only contain the markup of these annotation levels.

The multiple annotation approach provides a very elegant solution to questions that are of importance with regard to general annotation problems: (1) how to handle the problem of annotating overlapping hierarchies and (2) how to deal with heterogeneous tag sets.[1] Furthermore, most of the points of criticism can be rebutted. For example, the concerns regarding data consistency lose their bite with the advent of original editing tools for the creation of primary-data-identic annotation files. As a further example, much of the remainder of this article deals with the transformation from multiple annotation documents to an integrated representation format, i.e. one that is encoded within the constraints of the TEI tag set for feature structures. A point to be learned from this is that the advantages of other approaches can be married with the specific advantages of the multiple annotations approach by supplementing it with specific software tools.

Though most linguistic corpora to be archived by the sustainability project are already encoded in XML-based formats, they are still heterogeneous from a conceptual point of view. The majority of corpora are annotated using markup languages that are based on the Annotation Graph framework (Bird and Liberman, 2001), the upcoming Linguistic Annotation Format ISO standard (Ide, 2007), or according to several tag sets defined by or based upon the TEI guidelines. The GENAU approach comprises the separation of individual annotation layers contained in the original corpus data into multiple XML files, so that each file represents a single annotation layer only. Several automatic or semi-automatic tools and XSLT stylesheets have been developed to normalize and to transform the original data formats into multiple XML files (see Fig. 1).

The description of the GENAU approach given above focuses on the representation of the data within multiple XML files. A different perspective on markup technology directs the abstract model instead of the syntax of the annotations used. From that point of view, an XML document is a tree structure, i.e. a set of nodes connected by directed edges. The nodes in the tree represent XML elements, the leaves of this tree are the characters of which the text consists. All but one node of the tree must have a single parent. The node without a parent is called the root node. Of course, XML documents are only one of multiple ways to represent tree structures by means of a linear stream of text data. An alternative linearisation of trees is the labeled bracketing format often used in linguistics, e.g. (s (n mary) (vp (v supports) (np (det the) (n union)))).

The abstract model of the multiple XML files used by the GENAU approach is not a single tree but several trees. Since each of these trees spans the same leaves such a structure is called a multi-rooted tree. A multi-rooted tree has as many roots as annotation encoded layers. The multiple files used by the GENAU approach can be regarded as a linearisation of a multi-rooted tree. The next section describes an alternative approach to represent these structures.

## 3 The TEI Tag Set for the Annotation of Feature Structures as a Representation Format for Multi-rooted Trees

In addition to the encoding in multiple files, other representation formats can be used to represent multi-rooted trees. One of these formats is based on the TEI tag set for the representation of feature structures (Sperberg-McQueen and Burnard, 2001). Although this tag set was included in version P3 of the TEI Guidelines (Sperberg-McQueen and Burnard, 1994) and adopted as an ISO standard (ISO 24610-1:2006, 2006) in 2006, it is used only rarely in academic applications. This tag set allows for the merging of all annotation information into a single XML document instance—at the same time it enables us to mark up phenomena that are hard or almost impossible to annotate using conventional approaches. In many branches of formal linguistics, feature structures are a common representation format. For example, several variants of generative grammar are grounded on the descriptive device of feature structures and the most important operation defined upon them: unification.[2]

From a mathematical point of view, feature structures can be described as partial functions from sets of features (also: attributes) onto sets of values. The values can be atomic or complex. As complex values are feature structure themselves, feature structures can be nested. Another mathematical stance on feature structures is the directed acyclic graph perspective. Feature structures can be visualized straightforwardly in the form of attribute value matrices, see Fig. 2. Concerning the operation of unification, the result of the unification of two feature structures can be intuitively conceived of as the fusion of the information contained in both feature structures, if the respective information packages are compatible with each other. Witt *et al.* (2005) describe an application of unification for XML documents with concurrent markup.

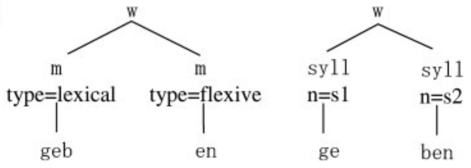
Since XML documents and feature structures are variants of directed acyclic graphs, there might be a straightforward mapping from one type of structural configuration onto the other. On closer investigation, however, some important differences can be uncovered. Sequential order, for example, plays an important role among the branches of subtrees of XML document trees, but it does not among the corresponding attribute-value pairs situated on an identical level within feature structures. Nevertheless, it is still possible to realize the desired mapping from the more restrictive to the less restrictive format using special representational means which have to be interpreted specifically.

The use of feature structures for the representation of multi-layered annotations is illustrated by means of a simplified example of a two tier annotation of a word. The German verb *geben* ('to give') is annotated morphologically and phonologically. The first annotation in (1)—or, correspondingly, the first tree structure in (2)—depicts the morphological annotation, the second one shows the phonological structure. Both annotations are marked up as single rooted trees.

(1)
```
<w>
    <m type="lexical">geb</m>
    <m type="flexive">en</m>
</w>

<w>
    <syll n="s1">ge</syll>
    <syll n="s2">ben</syll>
</w>
```

(2)



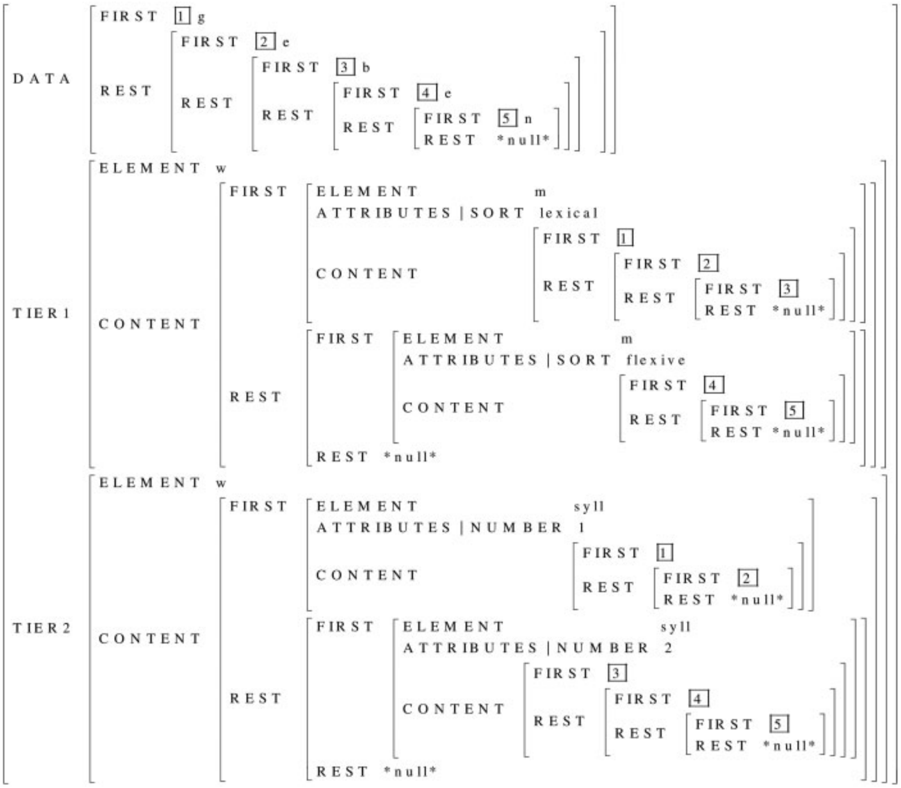


**Fig. 2** AVM representation of the annotation data example

Let us compare the attribute value matrix visualized in Fig. 2 to the tree representation format in (2): in order to express the information contained in both trees by means of a single feature structure, the concurring annotation layers could be embedded into the different top-level features, e.g. into the features TIER1 and TIER2, whereas the primary data are segmented into single indexed characters and represented along the remaining top-level feature DATA. Generally, sequential relations as those among the indexed characters under DATA are expressed by means of appropriate FIRST/REST value pair assignments that correspond to list notations. The solution for the representation of hierarchical relations consists in a similar mechanism: the exploitation of a special feature CONTENT which also embeds list-like feature structures such as those under DATA. Attributes are represented in a straightforward way by means of a mapping onto the value of the ATTRIBUTES feature. The anchoring of the annotation to the data is realized via a reference mechanism known as structure sharing or reentrancy, which is a commonplace among feature structures. It can be interpreted as token-identity and is indicated using co-indexed boxes here.

The TEI tag set for the representation of feature structures can be used to encode this feature structure in an XML-based format. Figure 3 shows the XML version of the attribute value notation depicted in Fig. 2.[3]

The backbone of the encoding consists in the use of `fs` elements for feature structures and `f` elements for features. From a conceptual point of view, this approach can be thought of as a 'retranslation' to XML. However, at the level of the automatic methods devised in order to realize the transformation into this exchange format, both steps (the transformation into a feature structure format and the

```
<fs>
 <f name="DATA">
   <fs>
     <f name="FIRST">
       <vLabel name="1">
         <symbol value="g"/>
       </vLabel>
     </f>
     <f name="REST">
       <fs>
         <f name="FIRST">
           <vLabel name="2">
             <symbol value="e"/>
           </vLabel>
         </f>
         <f name="REST">
           <fs>
             <f name="FIRST">
               <vLabel name="3">
                 <symbol value="b"/>
               </vLabel>
             </f>
             <f name="REST">
               <fs>
                 <f name="FIRST">
                   <vLabel name="4">
                     <symbol value="e"/>
                   </vLabel>
                 </f>
                 <f name="REST">
                   <fs>
                     <f name="FIRST">
                       <vLabel name="5">
                         <symbol value="n"/>
                       </vLabel>
                     </f>
                     <f name="REST">
                       <symbol value="*null*"/>
                     </f>
                   </fs>
                 </f>
               </fs>
             </f>
           </fs>
         </f>
       </fs>
     </f>
   </fs>
 </f>
 <f name="TIER1">
  ...
 </f>
 <f name="TIER2">
  ...
 </f>
</fs>
```

**Fig. 3** TEI-based feature structure representation of the AVM example

retranslation into XML) are broken down into a single step since the feature structure output can be directly represented as XML code that conforms to the TEI tag set standard.

The automatic methods to bring about the transformation consist in the subsequent execution of code written in Perl[4] and the application of XSLT stylesheet processing. The Perl code checks for the identity of the primary data among the multiple files corresponding to the different annotation layers to be integrated, while the XSLT stylesheet contains the actual transformation rules.

## 4 Comparison with Alternative Representation Formats

The list of possible alternatives to the use of TEI feature structures includes XCONCUR and the stand-off annotation approach of the NITE project. XCONCUR (Hilbert *et al.*, 2005, Schonefeld and Witt, 2006) can be characterized as a means of augmenting the XML standard with the optional CONCUR feature of the XML predecessor SGML—the syntax of XCONCUR is reminiscent of SGML with the CONCUR feature enabled. The basic mechanism is to prefix each element with an obligatory identity label for its respective annotation layer as conforming to this simple scheme: `(layer-id) name`. Here, of course, `layer-id` is a placeholder for the annotation layer label and name stands for the element's name. XCONCUR documents have to be well-formed. This condition is related to XML well-formedness via a projection to a set of well-formed XML documents. Each member of such a set can be conceptualized as representing the information content of a respective annotation layer. It is generated by way of decomposition from the original XCONCUR document, i.e. stripping the non-pertinent parts (see Witt *et al.*, 2007, also with respect to constraint-based cross-level validation).

The above example of a morphological and syllabic annotation of the German verb *geben* ('to give') can be represented in XCONCUR as follows:

```
(3) <?xconcur version="1.1" encoding="utf-8"?>
    <(l1)w>
    <(l2)w>
      <(l1)m type="lexical">
      <(l2)syll n="s1">
      ge
      </(l2)syll>
      <(l2)syll n="s2">
      b
      </(l1)m>
      <(l1)m type="flexive">
      en
      </(l2)syll>
      </(l1)m>
    <(l2)/w>
    <(l1)/w>
```

In comparison to the not even completely reproduced TEI-based representation in Fig. 3, this representation format is leaner. Obviously, this XCONCUR document is not a well-formed XML document due to the overlapping elements. The members of the projectable set, however, are in fact well-formed XML documents. Finally, like the TEI-based format this is also an integrative one, i.e. the whole information is packed into a single document. Both formats can be used as storage and exchange formats for multi-hierarchically annotated linguistic data.

The NITE project format exemplifies an XML-based approach to stand-off annotation. In particular, the format separates each coding for every observation into a separate file (Carletta *et al.*, 2003). A coding consists of one or more layers whose annotations can be arranged hierarchically as a tree structure.[5] For example, we may have separate phonological, morphological, syntactic and pragmatic codings for natural language data. With regard to the 'geben' example, we have simple morphological, syllabic and character codings. An observation consists of a piece of data to be annotated, e.g. a dialogue or, here, just a token of the verb 'geben'. The different coding files have to conform to the XML format. Links between them can be expressed via XLink/XPointer-mechanisms or according to an older, project-specific syntax that is also used in our example representation below.

```
(4) <root id="01.charachters">
      <charachter id="c_1" start="0" end="1"
        char="g"/>
```

```
    <charachter id="c_2" start="1" end="2"
      char="e"/>
    <charachter id="c_3" start="2" end="3"
      char="b"/>
    <charachter id="c_4" start="3" end="4"
      char="e"/>
    <charachter id="c_5" start="4" end="5"
      char="n"/>
  </root>
```

```
(5) <root id="01.syllabic">
    <w id="w_1">
    <syll id="s_1">
      <child href="01.charachters.xml#id('c_1')"/>
      <child href="01.charachters.xml#id('c_2')"/>
    </syll>
    <syll id="s_2">
      <child href="01.charachters.xml#id('c_3')"/>
      <child href="01.charachters.xml#id('c_4')"/>
      <child href="01.charachters.xml#id('c_5')"/>
    </syll>
    </w>
  </root>
```

```
(6) <root id="01.morphological">
    <w id="w_1">
     <m id="m_1" type="lexical">
      <child href="01.charachters.xml#id('c_1')"/>
      <child href="01.charachters.xml#id('c_2')"/>
      <child href="01.charachters.xml#id('c_3')"/>
     </m>
     <m id="m_2" type="flexive">
      <child href="01.charachters.xml#id('c_4')"/>
      <child href="01.charachters.xml#id('c_5')"/>
     </m>
    </w>
  </root>
```

The representation is separated into the three coding files listed as (4), (5), and (6). The names of these files are `01.charachter.xml`, `01.syllabic.xml`, and `01.morphological.xml`, respectively. The 01.-prefix binds the codings to the same observation piece. Annotations at the syllabic (5) and morphological (6) levels are grounded via a reference mechanism that exploits IDs that have been attached to elements at the 'foundational' character coding level (4).

Just as TEI-based feature structures, but unlike XCONCUR, the NITE project format uses XML and, therefore, inherits its advantages. However, unlike its two representation alternatives, the NITE format separates the information across different document instances and could therefore be criticized as being not integrative in a strict sense (at least in the sense of a narrow reading of that term). With regard to document length considerations, the NITE representation format seems to take a middle ground. On the one hand, it is not as lean as an XCONCUR representation, on the other, the NITE representations are not as lengthy as those produced by the TEI feature structure format.

## 5 Conclusions

We have shown that it is possible to use the TEI tag set for the representation of feature structures as a meta-representation format for linguistic annotation resources. The underlying architecture is described as well as the conceptual approach and issues in the transformation from multiple XML annotation files to single-file, XML-based, TEI-adherent feature structure representations.

The move to a feature structure meta-format is an interesting research question in its own right, since feature structures are such common representation formalism among linguists adhering to different grammar theories today. However, the ability to represent one's data in that format should not only create some level of interest among researchers familiar with the formalism—it might also open up new possibilities with regard to subsequent algorithmic processing developed against that background: Witt *et al.* (2005) demonstrates an example of 'crossing over' between classic themes in computational linguistics and new fields of application in text technology.

However, the use of TEI-based feature structure representations also has a disadvantage. As the short and rather simple examples above illustrate, the respective output documents tend to get fairly long and they are also somewhat more cumbersome to inspect manually. Hence, this format seems to be more appropriate as a storage and analysis format for machines to process rather than as a human oriented presentation format.

## Acknowledgements

The research presented in this article was funded by the German Research Foundation (DFG).

## References

**Bird, S. and Liberman, M.** (2001). A Formal Framework for Linguistic Annotation. *Speech Communication*, **33**(1–2): 23–60.

**Carletta, J., DeRose, S., Durusau, P., Piez, W., Sperberg-McQueen, C. M., Tennison, J., and Witt, A.** (2007). International Workshop on Markup of Overlapping Structures. In Usdin, B.T. (ed.), *Proceedings of Extreme Markup Languages 2007,* Montréal, Canada.

**Carletta, J., Kilgour, J., O'Donnell, T., Evert, S., and Voormann, H.** (2003). The NITE Object Model Library for Handling Structured Linguistic Annotation on Multimodal Data Sets. *Proceedings of the EACL Workshop on Language Technology and the Semantic Web (3rd Workshop on NLP and XML, NLPXML-2003)*. Language Technology and the Semantic Web. EACL-2003, Budapest. Association for Computational Linguistics.

**Carpenter, B.** (1992). *The Logic of Typed Feature Structures: With Applications to Unification Grammars, Logic Programs and Constraint Resolution*. Number 24 in Cambridge Tracts in Theoretical Computer Science. Cambridge: Cambridge University Press.

**DeRose, S.** (2004). Markup overlap: A Review and a Horse. In Usdin, B.T. (ed.), *Proceedings of Extreme Markup Languages 2004,* Montréal, Canada.

**Hilbert, M., Schonefeld, O., and Witt, A.** (2005). Making CONCUR Work. In Usdin, B.T. (ed.), *Proceedings of Extreme Markup Languages 2005,* Montréal, Canada.

**Ide, N.** (2007). Annotation Science: From Theory to Practice and Use. In Rehm, G., Witt, A., and Lemnitzer, L. (eds), *Data Structures for Linguistic Resources and Applications: Proceedings of the Bieenial GLDV Conference 2007*. Tübingen: Narr, pp. 3–7.

**ISO 24610-1:2006** (2006). *Language Resource Management – Feature Structures- Part 1: Feature Structure Representation*. International Organization for Standardization.

**Maas, J. F.** (2003). *NEXUS: Vollautomatische Konvertierung mehrfach XML-annotierter Texte in das NITE-XML Austauschformat*. Master's thesis, Bielefeld University.

**Rehm, G. and Witt, A.** (2007). Digital Text Resources for the Humanities: Legal Issues. Session Description. In Schmidt, S., Siemens, R., Kumar, A., and Unsworth, J. (eds), *Digital Humanities 2007*. Urbana-Champaign, USA: Graduate School of Library and Information Science, University of Illinois, Urbana-Champaign, pp. 161–162.

**Rehm, G., Eckart, R., Chiarcos, C., and Dellert, J.** (2008a). Ontology-Based XQuery'ing of XML-Encoded Language Resources on Multiple Annotation Layers. *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*. Marrakech, Morocco.

**Rehm, G., Schonefeld, O., Witt, A., Lehmberg, T., Chiarcos, C., Bechara, H., Eishold, F., Evang, K., Leshtanska, M., Savkov, A., and Stark, M.** (2008b). The Metadata-Database of a Next Generation Sustainability Web-Platform for Language Resources. *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*. Marrakech, Morocco.

**Rehm, G., Schonefeld, O., Witt, A., Hinrichs, E. and Reis, M.** (2009). Sustainability of Annotated Resources in Linguistics: A Web-Platform for Exploring, Querying, and Distributing Linguistic Corpora and Other Resources. *Literary and Linguistic Computing*, **24**(2): 193–210.

**Schonefeld, O. and Witt, A.** (2006). Towards validation of concurrent markup. In Usdin, B.T. (ed.), *Proceedings of Extreme Markup Languages 2006,* Montréal, Canada.

**Shieber, S.M.** (1986). *An Introduction to Unification-based Approaches to Grammar.* Number 4 in CSLI Lecture Notes. Stanford, CA: CSLI Publications.

**Sperberg-McQueen, C. M. and Burnard, L.** (1994). *TEI Guidelines for Electronic Text Encoding and Interchange (TEI P3)*, Chicago and Oxford: Text Encoding Initiative.

**Sperberg-McQueen, C. M. and Burnard, L.** (2001). *Guidelines for Electronic Text Encoding and Interchange (TEI P4)*. Chicago and Oxford: Text Encoding Initiative, chapter 16: Feature Structures.

**Witt, A.** (2004). Multiple Hierarchies: New Aspects of an Old Solution. In Usdin, B.T. (ed.), *Proceedings of Extreme Markup Languages 2004*, Montréal, Canada.

**Witt, A., Goecke, D. Sasaki, F., and Lüngen, H.** (2005). Unification of XML Documents with Concurrent

Markup. *Literary and Linguistic Computing*, **20**(1): 103–116.

**Witt, A., Schonefeld, O., Rehm, G., Khoo, J., and Evang, K.** (2007). On the Lossless Transformation of Single-File Multi-Layer Annotations into Multi-Rooted Trees. In Usdin, B.T. (ed.), *Proceedings of Extreme Markup Languages 2007*, Montréal, Canada.

**Wörner, K., Witt, A., Rehm, G., and Dipper, S.** (2006). Modelling Linguistic Data Structures. In Usdin, B.T. (ed.), *Proceedings of Extreme Markup Languages 2006*, Montréal, Canada.

## Notes

[1] Possible alternative solutions or workarounds to question (1) include those also mentioned in the TEI guidelines (CONCUR, milestone elements, fragmentation technique, and virtual joins) and, probably the most widely applied technique, stand-off annotation (see also Section 4). The namespace standard provides a possible solution to question (2), but not to question (1).

[2] Shieber (1986) gives an introduction on unification-based grammars, Carpenter (1992) provides the formal background on feature structures.

[3] Due to space restrictions, Fig. 3 only displays the representation of the first top-level feature of the attribute value matrix, i.e. the feature structure underneath DATA.

[4] Parts of the code are based on the NEXUS tool developed by Maas (2003).

[5] Relations among different codings and the shared data give rise to the multi-rooted tree perspective.