ISO/TC 37/SC 4 N522
Date: 2008-5-12
ISO/CD 24612
ISO/TC 37/SC 4/WG 1
Secretariat: KATS

Language resource management – Linguistic annotation framework

Warning

Document type: International standard

Document subtype:

Document stage: (30) Committee

Document language: E

iso_tc37_sc4_N522_wg1_CD_24612_LAF.doc

Copyright Manager

ISO Central Secretariat

1 rue de Varembé

1211 Geneva 20 Switzerland

tel. + 41 22 749 0111

fax + 41 22 749 0947

internet: iso@iso.org

Table of contents

Foreword

Introduction

1 Scope **7**

2 Normative references **7**

3 Terms and definitions **8**

4 LAF specification **10**

4.1 LAF data model *10*

4.2 LAF architecture *10*

4.2.1 Primary data 10

4.2.2 Annotation documents 11

4.2.3 References to primary data 11

4.2.4 Base segmentation **Error! Bookmark not defined.**

4.2.5 Primary data document header 12

4.2.6 Annotation document header 12

4.3 Pivot format *12*

4.3.1 GrAF XML schema type hierarchy **Error! Bookmark not defined.**

Annex A (Informative) GrAF XML schema **17**

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 24612 was prepared by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 4, *Language resource management*.

The main part of the document ISO 24612 consists of the following clauses under the general title *Language resource management — Linguistic annotation framework*:

- Scope
- Normative references
- Terms and definitions
- Overview
- LAF specification: data model, architecture, Pivot format, XML serializing, and the GrAF API
- Examples
- Merging annotations
- Using the graphs
- Deriving annotation schemes

The document is then followed by an informative Annex A (informative) GrAF XML schema.

Introduction

This standard provides a primary base for language resource management, developed by the Working Group 1 of ISO/TC 37/SC 4. The aim of the group focuses on basic descriptors.

Language resource management – Linguistic annotation framework

1 Scope

Effective creation, coding, processing and management of language resources is facilitated by a single high-level data model that supports analysis and design of both databases and intermediate formats. This International Standard specifies a model that has been designed for the purpose of providing guidance on the basic principles for representing linguistic resource annotation schemes.

This International Standard is designed to support the development and use of computer applications for linguistic annotations and representations and the exchange of such data between different applications. Standardization of data categories and methods for the specification of data structures are given in ISO 12620 and other related International Standards.

This International Standard specifies a framework designed to provide guidance on the basic principles for representing linguistic annotation schemes. This framework includes a meta-model and methods for describing specific annotation schemes expressed in XML. The mechanisms for implementing constraints in an annotation scheme are defined in this International Standard, but not the specific constraints for individual annotation schemes (which will be the subject of further parts of this standards).

This International Standard also defines the conditions that allow the data expressed in one annotation scheme to be mapped onto another annotation scheme and specifies a generic mapping tool (GMT), for this purpose.

2 Normative references

ISO/IEC 639, Information technology - ISO 639:1988, Code for the representation of names of languages.

ISO 639-2:1998, Code for the representation of names and languages-part 2: Alpha-3 code.

ISO/IEC 646:1991 Information technology - ISO 7-bit coded character set for information interchange.

ISO 3166-1:1997, Code for the representation of names of countries and their subdivisions - Part 1: Country codes

ISO 8601:1988, Data elements and interchange formats - Information interchange - Representation of dates and times.

ISO 8879:1986 (SGML) as extended by TC2 (ISO/IEC JTC 1/SC 34 N 029:1998-12-06) to allow for XML.

ISO/IEC 10646-1:1993, Information technology-Universal Multiple-Octet Coded Character Set (UCS)-Part 1: Architecture and basic multilingual plane.

ISO 24610 Feature structures -Part 1: Feature structure representation

ISO 12620 Computer applications in terminology - Data categories.

3 Terms and definitions

For the purposes of specific use in this standard, the following terms are defined as follows:

3.1

primary data

electronic representation of language

EXAMPLES: text, image, speech signal, etc.

NOTE: Typically, primary data objects are represented by "locations" in an electronic file, for example, the span of characters comprising a sentence or word, or a point at which a given temporal event begins or ends (as in speech annotation). More complex data objects may consist of a list or set of contiguous or non-contiguous locations in primary data.

3.2

annotation

process of adding linguistic information to **primary data** (3.1) or the linguistic information itself, independent of its representation

NOTE: the word "annotation" used as a verb refers to the addition of linguistic information to primary data, whereas its use as a noun refers to the linguistic information itself.

3.3

representation

format in which the **annotation** (3.2) is rendered, independent of its content

EXAMPLES: XML, list or bracketed format, tab-delimited text, etc.

NOTE: a phrase structure syntactic annotation and a dependency-based annotation may both be represented using XML, even though the annotation information itself is very different.

3.4

segmentation annotation

annotation (3.2) that delimits linguistic elements that appear in the **primary data** (3.1)

NOTE: these elements include: (1) continuous segments (appear contiguously in the primary data), (2) super- and sub-segments, where groups of segments will comprise the parts of a larger segment (e.g., contiguous word segments typically comprise a sentence segment), (3) discontinuous segments (linking continuous segments), and (4) landmarks (e.g time stamps) that note a point in the primary data. In current practice, segmental information may or may not appear in the document containing the primary data itself. Documents considered to be *read-only,* for example, might be segmented by specifying offsets into the primary document where a given segment begins and ends.

3.5

linguistic annotation

annotation (3.2) that provides linguistic information about the segments in the **primary data** (3.1)

EXAMPLE: A morpho-syntactic annotation in which a part of speech and lemma are associated with each segment in the data

NOTE: the identification of a segment as a word, sentence, noun phrase, etc. also constitutes linguistic annotation. In current practice, when it is possible to do so, segmentation and identification of the linguistic role or properties of that segment are often combined (e.g., syntactic bracketing, or delimiting each word in the document with an XML element that identifies the segment as a word, sentence, etc.).

3.6

stand-off annotation

annotation (3.2) layered over **primary data** (3.1) and instantiated in a document separate from that containing the **primary data** (3.1).

NOTE: stand-off annotations refer to specific locations in the primary data, by addressing character offsets, elements, etc. to which the annotation applies. Multiple stand-off annotation documents for a given type of annotation can refer to the same primary document (e.g., two different part of speech annotations for a given text). There is no requirement that a single XML-compliant document may be created by merging stand-off annotation documents with the primary data; that is, two annotation documents may specify trees over the primary data that contain overlapping hierarchies.

3.7

document

XML document containing **primary data** (3.1) or **annotations** (3.2)

NOTE: In LAF, documents containing textual data are encoded in UTF-8 or UTF-16 (default: UTF-8). Each document is associated with a header document that describes its contents, provenance, etc.; several primary data or annotation documents may be associated with the same header.

3.8

anchor

a fixed, immutable position in the artifact being annotated

NOTE: The medium determines how an anchor is described. For example, text anchors may be character offsets, audio anchors may be time offsets, video anchors may be time offsets or frame indices, image anchors may be coordinates.

3.9

region

the area of a document bounded by two or more anchors.

NOTE: Every region also defines *start* and *end* anchors, where the *start* anchor is the first anchor returned by an iterator and the *end* anchor is the last anchor returned by an iterator.

3.10

original

the artifact or **annotation** (3.2) in different format from which the **primary data** (3.1) is derived

4 LAF specification

LAF consists of the following:

- A data model for linguistic annotations and the data to which they apply.
- An architecture for representing language data and its annotations.
- An XML serialization of the data model, which describes the referential structure of annotations associated with language data, consisting of a directed graph or graphs whose terminal nodes (nodes with no outgoing edges) reference primary data. Vertices and edges may be labeled with *feature structures* describing linguistic properties of reachable nodes or spans of data. *n*-dimensional regions of primary data.
- An API for manipulating the data model.

4.1 LAF data model

The LAF data model consists of (1) a *referential structure* for associating stand-off annotations with primary data; and (2) a *feature structure representation* for annotation content. The data model for annotations thus comprises a directed graph referencing *n*-dimensional regions of primary data as well as other annotations, in which nodes are labeled with feature structures providing the annotation content. LAF conformance requires that an annotation scheme must be (or be rendered via the mapping) isomorphic to the LAF data model.

NOTE: LAF does not include specifications for annotation *content categories* (i.e., the contents of the associated linguistic phenomena).

4.2 LAF architecture

Language resources conforming to the LAF architecture consist of the following, which are described in more detail in the following section:

- One or more *primary data documents*;
- One or more documents referencing the primary data that may serve as the *base segmentation* for other annotations;
- Any number of additional *annotation documents* containing linguistic information that reference a base segmentation document or other annotation documents;
- A *header document* associated with each primary data document and annotation document (or a set of primary data or annotation documents treated as a logical unit).

It is recommended that when appropriate, each primary data document also be associated with an *original* document that contains the source from which the primary data was adapted or extracted for annotation (e.g., the original text in the file format of a particular word processor or file viewer).

4.2.1 Primary data

Primary data is frozen as "read-only" to preserve the integrity of references to locations within the document or documents. Corrections and modifications to the primary data should be treated as annotations and stored in a separate document.

In the general case, primary data does not contain markup of any kind. If markup does exist in primary data (e.g., HTML or XML tags), it is treated as a part of the data stream by referring

annotations; no distinction is made between markup and other characters in the data when referring to locations in the document. It is crucial to note that *LAF primary data documents are not XML documents*—they contain no markup that is interpreted during processing as anything other than a part of the primary data itself.

4.2.2 Annotation documents

Annotation documents contain information describing primary data. Annotation information may be directly associated with (linked to) primary data, or it may be associated with another annotation or annotations. In the latter case, the annotations are said to be *layered* over the primary data.

The granularity of the annotation—that is, the smallest information unit to which the annotation applies—is dependent on the application. A single annotation may cover a sentence, paragraph, document, or an entire corpus.

LAF recommends representing each of the annotation layers defined in ISO TC37 SC4 in separate annotation documents for the purposes of exchange.

4.2.3 References to primary data

Direct reference to locations in primary data is to virtual *nodes* located between each base unit of representation. For text, the default base unit is one character, as defined by ISO/IEC 10646-1 Information Technology – Universal Multiple-Octet Coded Character Set.

Nodes are numbered starting at 0, the position just prior to the first base unit in the document.

Location indexes are considered to fall between base units, starting at 0. For example, consider the text “My dog has fleas”:

```
                        1
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6
|M|y| |d|o|g| |h|a|s| |f|l|e|a|s|
```

The location indexes for each word are:

My : start=0, end=2
dog : start=3, end=6
has : start=7, end=10
fleas : start=11, end =16

Regions defined by location indexes need not be contiguous (i.e., there may be portions of the primary data not referenced by any annotation), but they should not, in general, overlap. Overlapping regions should be considered to be composed of finer-grained sub-components. For example, two spans <5, 9> and <7, 15> can be reconstrued as three spans a = <5, 7>, b = <7, 9>, and c = <9, 15>. Two graph nodes can then be created that reference nodes <a, b> and <b, c>, thereby providing the coverage of spans <5, 9> and <7, 15>.

Discontiguous regions are referenced by creating nodes for each of the components and adding a node that references them. References may include three or more nodes when referencing image or other non-textual data, or when dealing with discontiguous regions comprised of three or more components.

Regions in artifacts such as an image map or video may be defined in terms of anchors specifying one or more coordinates, frame indexes, etc. The media types included in the corpus are defined in the corpus header. The corpus header also associates an anchor type each media type. The header for each primary data document identifies the medium for that document, which in turn identifies the type of anchors used to identify regions in that document.

4.3 Headers

LAF defines a header for a corpus consisting of a collection of primary data documents and annotations; as well as headers for primary data and annotation documents themselves. This set of headers provides all metadata describing the provenance and encoding conventions for the data and its annotations; information required for processing such as anchor types; and relations among primary data and annotation documents in the corpus.

4.3.1 Corpus header

The corpus header describes the corpus as a whole, including its contents, file structure, and encoding; and establishes definitions that are used in the primary data document and annotation documents headers. Among these are:

- *Categories* used to describe primary data documents, typically domain/subject area for general text.
- *File types* providing their naming conventions, media, annotation type, and dependencies (i.e., other file types that are referenced and therefore required).
- *Annotation sets* describing the annotations in the corpus, including their names, the layer to which they belong, creator, and a link to relevant documentation.
- *Media definitions* specifying the media types included in the corpus and file naming conventions for files containing data of that type.
- *Anchor types* associating anchor type definitions with media types.
- *Layer definitions* providing the names, descriptions, and dependencies among annotation layers.

4.3.2 Primary data document header

Each primary data document is associated with a header file containing information describing it contents. Because the primary data document is not an XML document, the LAF primary data header is obligatorily provided as a standalone header file.

The primary data document header provides information about the source and contents of the primary data, and specifies category definitions and medium type by reference to definitions in the corpus header.

The primary data document header provides the PID for the primary data document and all associated annotation documents. This information is used for processing the corpus.

4.3.3 Annotation document header

The LAF annotation document header includes a relevant subset of elements from the primary data header (i.e., those that describe the file contents rather than the provenance of an original text, etc.), together with additional elements that provide or point to information concerning the annotation content categories and dependencies between the annotation document and other documents. The annotation document header is not a separate document, but rather is included at the beginning of the annotation document.

4.4 Pivot format

LAF provides an XML serialization of the data model that is designated as the *pivot format*. Although the LAF pivot format may be used in any context, it is assumed that users will represent annotations using their own formats, which may then be transduced to the LAF pivot format for the purposes of exchange, merging, and comparison.

The Graph Annotation Format (GrAF) specifies the XML serialization of the LAF pivot format.

In GrAF:

The fundamental data structure is a graph consisting of nodes and edges.

An annotation is a label and a feature structure associated with a node or an edge in the graph.

A feature structure is a list of features or nested feature structures.

A feature is a mapping from one string (the name) to another string (the value).

Nodes may be associated with regions in the primary document defined in a base segmentation document, or related to other nodes in the same or another annotation document by one or more edges.

An edge represents a relationship between nodes. By default edges from a node represent an ordered set of constituents of the annotation associated with the node. Other relationships may be specified by associating an annotation with the edge.

The GrAF model is similar to most other document models with the following exceptions:

- Nodes in the graph do not represent annotations; they are simply place holders that may be associated with zero or more annotations.
- Edges in the graph are first class citizens of the data model. In many data models the edges between annotations are implied by the nesting of tags (XML, Lisp) or by listing children by reference (W3C DOM, UIMA). In GrAF the edges between annotations are explicitly represented as objects and may also be annotated.

The GrAF XML schema for annotation documents is included in Annex A.

4.5 XML elements for annotation documents

The overall structure of XML elements in a GrAF annotation file is given in Figure 1.

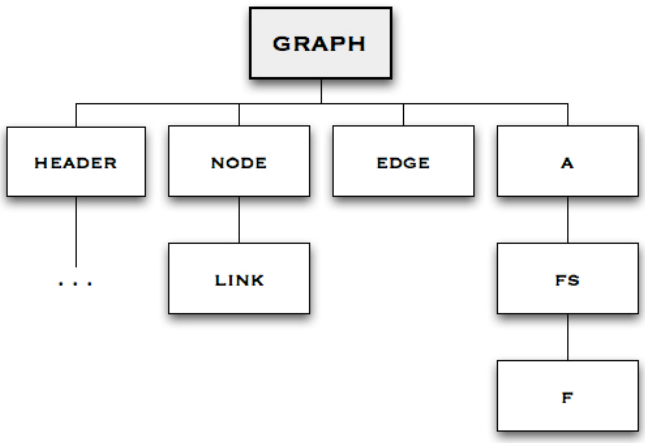

Figure 1. Parent-child relations of top-level XML elements in a GrAF annotation document.

The root node of the document is defined as follows:

<graph> The root node of the graph.

Attribute: *xmlns* : namespace declaration for the GrAF schema.

Example: `<graph xmlns="http://www.xces.org/ns/GrAF/1.0/">`

4.5.1 The header

The XML element structure for the header is given in Figure 2.

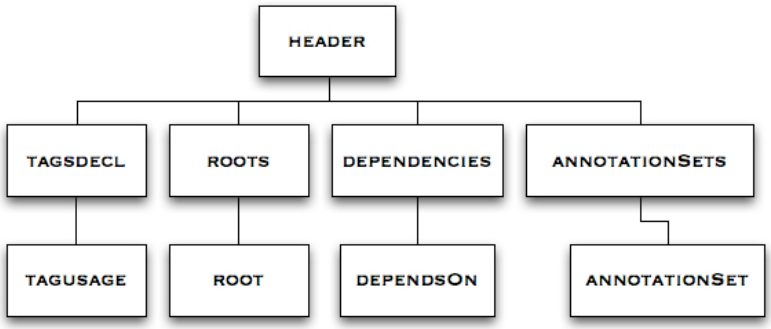

Figure 2. XML elements in the header of a GrAF annotation document.

<header> Bracketing tag for elements of the annotation document header. No attributes.

Header elements are as follows:

<tagsDecl>[1] Provides a list of the annotation labels used in the document and their frequencies. No attributes.

<tagUsage> Information for individual annotation categories.

Attributes:

gi [string] element name

occurs [integer] number of occurrences in the document

<roots> (optional) Groups a list of one or more root nodes of the graph or graphs contained in the document, where applicable, for use in processing. No attributes.

<root> Provides the ID of a root node as tag content. No attributes.[2]

<dependencies>Groups the list of document types required to process the annotations in this document. No attributes.

<dependsOn> Identifies a required file type, as defined in the corpus header. No attributes.

<annotationSets>Groups the list of annotation sets used in this document. No attributes.

---

[1] This has to be changed to avoid confusion with the TEI tag of the same name, perhaps to catDecl and catUsage.

[2] In future LAF versions this will be changed to an empty tag with a *ref* attribute, whose value provides the root node ID, for consistency.

<annotationSet>Gives the name and type of an annotation set, as defined in the corpus header.

Attributes:

name [string] The declared name of the annotation set as specified in the corpus header.

type [URI] Unique identifier used for the annotation set. Typically it is the URI of the annotation scheme documentation or project site.

4.5.2 Graph and annotation elements

<node> A node in the graph. Attribute: *xml:id*. The <node> element is empty when it is connected by an <edge> element to another node in the graph (i.e., when the node is a non-terminal node). A child <link> element is used when the node refers to a region or regions of primary data (i.e., when the node is a terminal/leaf node).

<link> Provides the IDs of segment(s) in a base segmentation document when the node is associated with one or more regions of primary data.

Attribute:

targets [list of one or more space-delimited segment IDs].

Note that the base segmentation document in which the segments referenced appear is specified in the header.

Example:
```
<node xml:id="penn-n10">
   <link targets="seg-r20"/>
</node>
```

<edge> An edge in the graph. Attributes: *xml:id, from* [node ID]*, to* [node ID].

Example: `<edge xml:id="e595" from="fn-n361" to="fn-n348"/>`

<a> Provides annotation information associated with a node or edge. This tag may be empty if the annotation consists of a label only.

Attributes:

label [string] : Provides the label of the annotation. This may be the string used to identify the annotation as described by the annotation documentation, a category identifier from a data category registry, an identifier from a feature structure library, or any reference to an external annotation specification.

ref [node or edge ID] : The ID of the node or edge with which the annotation is associated.

as [string] : The name of the annotation set of which this annotation is a part, as specified in the corpus header.

<fs> A feature structure providing additional annotation information. No attributes. The <a> element may contain one or more <fs> elements. The <fs> element may contain one or more <f> elements.

<f> Provides an attribute/value pair. In the concise form specified for GrAF annotations[3], the <f> element is empty and includes attributes providing simple name/value pairs.

Attributes:

name [string] : the name of the attribute, as specified in the documentation of the annotation scheme.

value [string] : the value for the attribute, taken from possible values provided in the documentation of the annotation scheme.

Example:

```
<a label="FE" ref="fn-n361" as="FrameNet">
  <fs>
    <f name="FE" value="Perceiver"/>
    <f name="rank" value="1"/>
    <f name="GF" value="Ext"/>
    <f name="PT" value="NP"/>
  </fs>
</a>
```

More complex feature structures may be represented according to the specification in ISO 24610-1, *Language resource management — Feature structures — Part 1: Feature structure representation*, which should be consulted for details.

4.6 Example

```
<graph xmlns="http://www.xces.org/ns/GrAF/1.0/">
    <header>
        <tagsDecl>
            <tagUsage gi="fullTextAnnotation" occurs="1"/>
            <tagUsage gi="Target" occurs="171"/>
            <tagUsage gi="FE" occurs="372"/>
            <tagUsage gi="sentence" occurs="32"/>
            <tagUsage gi="annotationSet" occurs="171"/>
            <tagUsage gi="NamedEntity" occurs="32"/>
        </tagsDecl>
        <roots>
            <root>fn-n0</root>
        </roots>
        <dependencies>
            <dependsOn type="fntok"/>
        </dependencies>
        <annotationSets>
            <annotationSet name="FrameNet"
                  type="http://framenet.icsi.berkeley.edu"/>
        </annotationSets>
    </header>
    …
    <node xml:id="fn-n156"/>
    <a label="FE" ref="fn-n156" as="FrameNet">
        <fs>
            <f name="FE" value="Speaker"/>
            <f name="rank" value="1"/>
            <f name="GF" value="Ext"/>
            <f name="PT" value="NP"/>
        </fs>
```

[3] Note that the GrAF concise form differs from the ISO specification for the <f> (feature) element in ISO 24610-1, in that it allows a *value* attribute that can take any string value instead of an *fVal* attribute that takes a pointer to an externally-defined data category as its value.

```
      </a>
      <edge xml:id="e233" from="fn-n156" to="fn-n133"/>
      <edge xml:id="e232" from="fn-n155" to="fn-n156"/>
      <edge xml:id="e231" from="fn-n132" to="fn-n155"/>
```

Annex A (Informative) GrAF XML schema for annotation documents

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified"
targetNamespace="http://www.xces.org/ns/GrAF/0.99/"
xmlns:graf="http://www.xces.org/ns/GrAF/0.99/">
  <xs:import namespace="http://www.w3.org/XML/1998/namespace" schemaLocation="xml.xsd"/>
  <!--
    Schema generated from ODD source 2010-02-03T12:02:14Z.
    TEI Edition: 1.3.0. Last updated on February 1st 2009.
  -->
  <!--

    All rights reserved. It is permitted to download this electronic file, to
    make a copy and to print out the content for the purpose of preparing ISO
    and IEC documents only. You may not copy or mirror the file, or any part of
    it, for any other purpose without permission in writing from the publishers.
  -->
  <!--
    Schema generated from ODD source 2009-06-21T11:06:38Z.
    TEI Edition: 1.4.0. Last updated on June 20th 2009.

  -->
  <!---->
  <xs:attributeGroup name="att.global.attributes">
    <xs:attributeGroup ref="graf:att.global.attribute.xmlid"/>
    <xs:attributeGroup ref="graf:att.global.attribute.n"/>
    <xs:attributeGroup ref="graf:att.global.attribute.xmllang"/>
    <xs:attributeGroup ref="graf:att.global.attribute.rend"/>
    <xs:attributeGroup ref="graf:att.global.attribute.rendition"/>
    <xs:attributeGroup ref="graf:att.global.attribute.xmlbase"/>
  </xs:attributeGroup>
  <xs:attributeGroup name="att.global.attribute.required.xmlid">
    <xs:attribute ref="xml:id" use="required"/>
  </xs:attributeGroup>
  <xs:attributeGroup name="att.global.attribute.xmlid">
    <xs:attribute ref="xml:id"/>
  </xs:attributeGroup>
  <xs:attributeGroup name="att.global.attribute.n">
    <xs:attribute name="n">
      <xs:annotation>
        <xs:documentation>(number) gives a number (or other label) for an element, which is
not necessarily unique within
        the document.</xs:documentation>
      </xs:annotation>
      <xs:simpleType>
        <xs:restriction>
          <xs:simpleType>
            <xs:list>
              <xs:simpleType>
                <xs:restriction base="xs:token">
                  <xs:pattern value="(\p{L}|\p{N}|\p{P}|\p{S})+"/>
                </xs:restriction>
              </xs:simpleType>
            </xs:list>
          </xs:simpleType>
          <xs:minLength value="1"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
  </xs:attributeGroup>
  <xs:attributeGroup name="att.global.attribute.xmllang">
    <xs:attribute ref="xml:lang"/>
  </xs:attributeGroup>
  <xs:attributeGroup name="att.global.attribute.rend">
    <xs:attribute name="rend">
      <xs:annotation>
        <xs:documentation>(rendition) indicates how the element in question was rendered or
presented in the source text.</xs:documentation>
      </xs:annotation>
      <xs:simpleType>
        <xs:restriction>
```

```
          <xs:simpleType>
            <xs:list>
              <xs:simpleType>
                <xs:restriction base="xs:token">
                  <xs:pattern value="(\p{L}|\p{N}|\p{P}|\p{S})+"/>
                </xs:restriction>
              </xs:simpleType>
            </xs:list>
          </xs:simpleType>
          <xs:minLength value="1"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
  </xs:attributeGroup>
  <xs:attributeGroup name="att.global.attribute.rendition">
    <xs:attribute name="rendition">
      <xs:annotation>
        <xs:documentation>points to a description of the rendering or presentation used for
this element in the
        source text.</xs:documentation>
      </xs:annotation>
      <xs:simpleType>
        <xs:restriction>
          <xs:simpleType>
            <xs:list itemType="xs:anyURI"/>
          </xs:simpleType>
          <xs:minLength value="1"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
  </xs:attributeGroup>
  <xs:attributeGroup name="att.global.attribute.xmlbase">
    <xs:attribute ref="xml:base"/>
  </xs:attributeGroup>
  <xs:group name="model.featureVal.complex">
    <xs:choice>
      <xs:element ref="graf:fs"/>
      <xs:element ref="graf:vColl"/>
      <xs:element ref="graf:vNot"/>
      <xs:element ref="graf:vMerge"/>
    </xs:choice>
  </xs:group>
  <xs:group name="model.featureVal.single">
    <xs:choice>
      <xs:element ref="graf:binary"/>
      <xs:element ref="graf:symbol"/>
      <xs:element ref="graf:numeric"/>
      <xs:element ref="graf:string"/>
      <xs:element ref="graf:vLabel"/>
      <xs:element ref="graf:default"/>
      <xs:element ref="graf:vAlt"/>
    </xs:choice>
  </xs:group>
  <xs:group name="model.featureVal">
    <xs:choice>
      <xs:group ref="graf:model.featureVal.complex"/>
      <xs:group ref="graf:model.featureVal.single"/>
    </xs:choice>
  </xs:group>
  <xs:element name="fsDecl">
    <xs:annotation>
      <xs:documentation>(feature structure declaration) declares one type of feature
structure.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element minOccurs="0" ref="graf:fsDescr"/>
        <xs:element maxOccurs="unbounded" ref="graf:fDecl"/>
        <xs:element minOccurs="0" ref="graf:fsConstraints"/>
      </xs:sequence>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
      <xs:attribute name="type" use="required" type="xs:Name">
        <xs:annotation>
          <xs:documentation>gives a name for the type of feature structure being
declared.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
      <xs:attribute name="baseTypes">
        <xs:annotation>
```

```
          <xs:documentation>gives the name of one or more typed feature structures
        from which this type inherits feature specifications and
        constraints;
        if this type includes a feature specification
        with the same name as that of any of those specified by this
        attribute, or if more than one specification of the same name
        is inherited, then the set of possible values is defined by
        unification. Similarly, the set of constraints applicable is
        derived by combining those specified explicitly within this
        element with those implied by the baseTypes
        attribute. When no baseTypes attribute is specified, no
        feature specification or constraint is inherited.

      </xs:documentation>
        </xs:annotation>
        <xs:simpleType>
          <xs:restriction>
            <xs:simpleType>
              <xs:list itemType="xs:Name"/>
            </xs:simpleType>
            <xs:minLength value="1"/>
          </xs:restriction>
        </xs:simpleType>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="fsDescr">
    <xs:annotation>
      <xs:documentation>(feature system description (in FSD)) describes in prose what is
represented by the type of feature
structure declared in the enclosing fsDecl.</xs:documentation>
    </xs:annotation>
    <xs:complexType mixed="true">
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="fsdLink">
    <xs:annotation>
      <xs:documentation>(feature structure declaration link) associates the name of a typed
feature structure with a feature
structure declaration for it.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
      <xs:attribute name="type" use="required" type="xs:Name">
        <xs:annotation>
          <xs:documentation>identifies the type of feature structure to be documented;
        this will be the value of the type attribute on at least one
feature structure.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
      <xs:attribute name="target" use="required" type="xs:anyURI">
        <xs:annotation>
          <xs:documentation>supplies a pointer to a feature structure declaration
        (fsDecl) element within the current document or elsewhere. </xs:documentation>
        </xs:annotation>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="fDecl">
    <xs:annotation>
      <xs:documentation>(feature declaration) declares a single feature, specifying its name,
organization,
range of allowed values, and optionally its default value.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element minOccurs="0" ref="graf:fDescr"/>
        <xs:element ref="graf:vRange"/>
        <xs:element minOccurs="0" ref="graf:vDefault"/>
      </xs:sequence>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
      <xs:attribute name="name" use="required" type="xs:Name">
        <xs:annotation>
          <xs:documentation>indicates the name of the feature being declared; matches the
name attribute of f elements in the text.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
      <xs:attribute name="optional" default="true" type="xs:boolean">
```

```
        <xs:annotation>
          <xs:documentation>indicates whether or not the value of this feature may
        be present.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="fDescr">
    <xs:annotation>
      <xs:documentation>(feature description (in FSD)) describes in prose what is represented
by the feature being
declared and its values.</xs:documentation>
    </xs:annotation>
    <xs:complexType mixed="true">
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="vRange">
    <xs:annotation>
      <xs:documentation>(value range) defines the range of allowed values for a feature, in
the form of
an fs, vAlt, or primitive value;
for the value of an f to be valid, it must be
subsumed by the specified range; if the f
contains multiple values (as sanctioned by the org attribute),
then each value must be subsumed by the vRange.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:group ref="graf:model.featureVal"/>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="vDefault">
    <xs:annotation>
      <xs:documentation>(value default) declares the default value to be supplied when a
feature structure
does not contain an instance of f for this name; if
unconditional, it is specified as one (or, depending on the value of
the org attribute of the enclosing fDecl) more
fs elements or primitive values;
if conditional, it is specified as
one or more if elements; if no default is specified, or no
condition matches, the value none is assumed.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:choice>
        <xs:group maxOccurs="unbounded" ref="graf:model.featureVal"/>
        <xs:element maxOccurs="unbounded" ref="graf:if"/>
      </xs:choice>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="if">
    <xs:annotation>
      <xs:documentation>defines a conditional default value for a feature; the condition
is specified as a feature structure, and is met if it
subsumes the feature structure in the text for which a
default value is sought.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:choice>
          <xs:element ref="graf:fs"/>
          <xs:element ref="graf:f"/>
        </xs:choice>
        <xs:element ref="graf:then"/>
        <xs:group ref="graf:model.featureVal"/>
      </xs:sequence>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="then">
    <xs:annotation>
      <xs:documentation>separates the condition from the default in an if, or
the antecedent and the consequent in a cond element.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
```

```
  </xs:element>
  <xs:element name="fsConstraints">
    <xs:annotation>
      <xs:documentation>(feature-structure constraints) specifies constraints on the content
of valid feature
structures.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:choice minOccurs="0" maxOccurs="unbounded">
        <xs:element ref="graf:cond"/>
        <xs:element ref="graf:bicond"/>
      </xs:choice>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="cond">
    <xs:annotation>
      <xs:documentation>(conditional feature-structure constraint) defines a conditional
feature-structure constraint; the consequent
and the antecedent are specified as feature structures or
feature-structure collections; the constraint is satisfied if both the
antecedent and the consequent subsume a given feature
structure, or if the antecedent does not.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:choice>
          <xs:element ref="graf:fs"/>
          <xs:element ref="graf:f"/>
        </xs:choice>
        <xs:element ref="graf:then"/>
        <xs:choice>
          <xs:element ref="graf:fs"/>
          <xs:element ref="graf:f"/>
        </xs:choice>
      </xs:sequence>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="bicond">
    <xs:annotation>
      <xs:documentation>(bi-conditional feature-structure constraint) defines a biconditional
feature-structure constraint; both
consequent and antecedent are specified as feature structures or groups
of feature structures; the constraint is satisfied if both
subsume a given feature structure, or if both do not.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:choice>
          <xs:element ref="graf:fs"/>
          <xs:element ref="graf:f"/>
        </xs:choice>
        <xs:element ref="graf:iff"/>
        <xs:choice>
          <xs:element ref="graf:fs"/>
          <xs:element ref="graf:f"/>
        </xs:choice>
      </xs:sequence>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="iff">
    <xs:annotation>
      <xs:documentation>(if and only if) separates the condition from the consequence in a
bicond
element.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="fs">
    <xs:annotation>
      <xs:documentation>(feature structure) represents a feature structure, that is, a
  collection of feature-value pairs organized as a
structural unit.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
```

```
      <xs:sequence>
        <xs:element minOccurs="0" maxOccurs="unbounded" ref="graf:f"/>
      </xs:sequence>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
      <xs:attribute name="type" type="xs:Name">
        <xs:annotation>
          <xs:documentation>specifies the type of the feature structure.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
      <xs:attribute name="feats">
        <xs:annotation>
          <xs:documentation>(features) references the feature-value specifications making up
this feature structure.</xs:documentation>
        </xs:annotation>
        <xs:simpleType>
          <xs:restriction>
            <xs:simpleType>
              <xs:list itemType="xs:anyURI"/>
            </xs:simpleType>
            <xs:minLength value="1"/>
          </xs:restriction>
        </xs:simpleType>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="f">
    <xs:annotation>
      <xs:documentation>(feature) represents a feature value specification, that
  is, the association of a name with a value of any of several different
types.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:group minOccurs="0" maxOccurs="unbounded" ref="graf:model.featureVal"/>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
      <xs:attribute name="name" use="required" type="xs:Name">
        <xs:annotation>
          <xs:documentation>provides a name for the feature.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
      <xs:attribute name="fVal" type="xs:anyURI">
        <xs:annotation>
          <xs:documentation>(feature value) references any element which can be used to
represent the
  value of a feature.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
      <xs:attribute name="value" type="xs:string">
        <xs:annotation>
          <xs:documentation>feature value of an unknown type.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="binary">
    <xs:annotation>
      <xs:documentation>(binary value) represents the value part of a feature-value
specification which can contain either
  of exactly two possible values.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
      <xs:attribute name="value" use="required" type="xs:boolean">
        <xs:annotation>
          <xs:documentation>supplies a binary value.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="symbol">
    <xs:annotation>
      <xs:documentation>(symbolic value) represents the value part of a feature-value
specification
  which contains one of a finite list of symbols.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
      <xs:attribute name="value" use="required">
        <xs:annotation>
          <xs:documentation>supplies the symbolic value for the feature, one of a finite list
```

```
that
may be specified in a feature declaration.</xs:documentation>
        </xs:annotation>
        <xs:simpleType>
          <xs:restriction base="xs:token">
            <xs:pattern value="(\p{L}|\p{N}|\p{P}|\p{S})+"/>
          </xs:restriction>
        </xs:simpleType>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="numeric">
    <xs:annotation>
      <xs:documentation>(numeric value) represents the value part of a feature-value
specification
  which contains a numeric value or range.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
      <xs:attribute name="value" use="required">
        <xs:annotation>
          <xs:documentation>supplies a lower bound for the numeric value represented,
      and also (if max is not supplied) its upper bound.</xs:documentation>
        </xs:annotation>
        <xs:simpleType>
          <xs:union memberTypes="xs:double xs:decimal">
            <xs:simpleType>
              <xs:restriction base="xs:token">
                <xs:pattern value="(\-?[\d]+/\-?[\d]+)"/>
              </xs:restriction>
            </xs:simpleType>
          </xs:union>
        </xs:simpleType>
      </xs:attribute>
      <xs:attribute name="max">
        <xs:annotation>
          <xs:documentation>supplies an upper bound for the numeric value
represented.</xs:documentation>
        </xs:annotation>
        <xs:simpleType>
          <xs:union memberTypes="xs:double xs:decimal">
            <xs:simpleType>
              <xs:restriction base="xs:token">
                <xs:pattern value="(\-?[\d]+/\-?[\d]+)"/>
              </xs:restriction>
            </xs:simpleType>
          </xs:union>
        </xs:simpleType>
      </xs:attribute>
      <xs:attribute name="trunc" type="xs:boolean">
        <xs:annotation>
          <xs:documentation>specifies whether the value represented should be
         truncated to give an integer value.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="string">
    <xs:annotation>
      <xs:documentation>(string value) represents the value part of a feature-value
specification
  which contains a string.</xs:documentation>
    </xs:annotation>
    <xs:complexType mixed="true">
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="vLabel">
    <xs:annotation>
      <xs:documentation>(value label) represents the value part of a feature-value
specification
  which appears at more than one point in a feature structure.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:group minOccurs="0" ref="graf:model.featureVal"/>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
      <xs:attribute name="name" use="required">
        <xs:annotation>
          <xs:documentation>supplies a name for the sharing point.</xs:documentation>
```

```
        </xs:annotation>
        <xs:simpleType>
          <xs:restriction base="xs:token">
            <xs:pattern value="(\p{L}|\p{N}|\p{P}|\p{S})+"/>
          </xs:restriction>
        </xs:simpleType>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="vColl">
    <xs:annotation>
      <xs:documentation>(collection of values) represents the value part of a feature-value
specification
  which contains multiple values organized as a set, bag, or list.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:choice minOccurs="0" maxOccurs="unbounded">
        <xs:element ref="graf:fs"/>
        <xs:group ref="graf:model.featureVal.single"/>
      </xs:choice>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
      <xs:attribute name="org">
        <xs:annotation>
          <xs:documentation>(organization) indicates organization of given value or values as
set, bag or list.</xs:documentation>
        </xs:annotation>
        <xs:simpleType>
          <xs:restriction base="xs:token">
            <xs:enumeration value="set">
              <xs:annotation>
                <xs:documentation>indicates that the given values are organized as a set.
   </xs:documentation>
              </xs:annotation>
            </xs:enumeration>
            <xs:enumeration value="bag">
              <xs:annotation>
                <xs:documentation>indicates that the given values are organized as a
bag (multiset).
   </xs:documentation>
              </xs:annotation>
            </xs:enumeration>
            <xs:enumeration value="list">
              <xs:annotation>
                <xs:documentation>indicates that the given values are organized as a
list.</xs:documentation>
              </xs:annotation>
            </xs:enumeration>
          </xs:restriction>
        </xs:simpleType>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="default">
    <xs:annotation>
      <xs:documentation>(default feature value) represents the value part of a feature-value
specification
  which contains a defaulted value.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="vAlt">
    <xs:annotation>
      <xs:documentation>(value alternation) represents the value part of a feature-value
specification
  which contains a set of values, only one of which can be valid.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:group ref="graf:model.featureVal"/>
        <xs:group maxOccurs="unbounded" ref="graf:model.featureVal"/>
      </xs:sequence>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="vNot">
    <xs:annotation>
      <xs:documentation>(value negation) represents a feature value which is the negation of
```

```
its content.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:group ref="graf:model.featureVal"/>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="vMerge">
    <xs:annotation>
      <xs:documentation>(merged collection of values) represents a feature value which is the
result of merging
  together the feature values contained by its children, using the organization
  specified by the org attribute.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:group maxOccurs="unbounded" ref="graf:model.featureVal"/>
      <xs:attributeGroup ref="graf:att.global.attributes"/>
      <xs:attribute name="org">
        <xs:annotation>
          <xs:documentation>indicates the organization of the resulting merged values as  set,
bag or list.</xs:documentation>
        </xs:annotation>
        <xs:simpleType>
          <xs:restriction base="xs:token">
            <xs:enumeration value="set">
              <xs:annotation>
                <xs:documentation>indicates that the resulting values are organized as a
set.</xs:documentation>
              </xs:annotation>
            </xs:enumeration>
            <xs:enumeration value="bag">
              <xs:annotation>
                <xs:documentation>indicates that the resulting values are organized as a bag
(multiset).
   </xs:documentation>
              </xs:annotation>
            </xs:enumeration>
            <xs:enumeration value="list">
              <xs:annotation>
                <xs:documentation>indicates that the resulting values are organized as a
list.</xs:documentation>
              </xs:annotation>
            </xs:enumeration>
          </xs:restriction>
        </xs:simpleType>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:attributeGroup name="graf.att.id.required.attributes">
    <xs:attributeGroup ref="graf:att.global.attribute.required.xmlid"/>
  </xs:attributeGroup>
  <xs:attributeGroup name="graf.att.id.optional.attributes">
    <xs:attributeGroup ref="graf:att.global.attribute.xmlid"/>
  </xs:attributeGroup>
  <xs:attributeGroup name="graf.att.type.attributes">
    <xs:attributeGroup ref="graf:graf.att.type.attribute.type"/>
  </xs:attributeGroup>
  <xs:attributeGroup name="graf.att.type.attribute.type">
    <xs:attribute name="type" type="xs:string">
      <xs:annotation>
        <xs:documentation/>
      </xs:annotation>
    </xs:attribute>
  </xs:attributeGroup>
  <xs:attributeGroup name="graf.att.location.attributes">
    <xs:attributeGroup ref="graf:graf.att.location.attribute.loc"/>
  </xs:attributeGroup>
  <xs:attributeGroup name="graf.att.location.attribute.loc">
    <xs:attribute name="loc" type="xs:anyURI">
      <xs:annotation>
        <xs:documentation>The loc (location) attribute is used to specify
                                    the location of an external resource. All locations are
specified
                               by Unique Resource Locators (URL).</xs:documentation>
      </xs:annotation>
    </xs:attribute>
  </xs:attributeGroup>
  <xs:attributeGroup name="graf.att.name.attributes">
    <xs:attributeGroup ref="graf:graf.att.name.attribute.name"/>
```

```
  </xs:attributeGroup>
  <xs:attributeGroup name="graf.att.name.attribute.name">
    <xs:attribute name="name" type="xs:string">
      <xs:annotation>
        <xs:documentation/>
      </xs:annotation>
    </xs:attribute>
  </xs:attributeGroup>
  <xs:group name="graf_graf.class.graph">
    <xs:choice>
      <xs:element ref="graf:node"/>
      <xs:element ref="graf:edge"/>
      <xs:element ref="graf:region"/>
      <xs:element ref="graf:a"/>
      <xs:element ref="graf:as"/>
    </xs:choice>
  </xs:group>
  <xs:element name="graph">
    <xs:annotation>
      <xs:documentation>The root element of a GrAF document. Contains the
                        specification for a single graph object.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="graf:header"/>
        <xs:element minOccurs="0" ref="graf:fs"/>
        <xs:group minOccurs="0" maxOccurs="unbounded" ref="graf:graf_graf.class.graph"/>
      </xs:sequence>
      <xs:attributeGroup ref="graf:graf.att.id.optional.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="node">
    <xs:annotation>
      <xs:documentation>Represents a node (vertex) in the graph.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element minOccurs="0" maxOccurs="unbounded" ref="graf:link"/>
      </xs:sequence>
      <xs:attributeGroup ref="graf:graf.att.id.required.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="edge">
    <xs:annotation>
      <xs:documentation>An edge (arc) between two nodes.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:attributeGroup ref="graf:graf.att.id.optional.attributes"/>
      <xs:attribute name="from" use="required">
        <xs:annotation>
          <xs:documentation/>
        </xs:annotation>
      </xs:attribute>
      <xs:attribute name="to" use="required">
        <xs:annotation>
          <xs:documentation/>
        </xs:annotation>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="region">
    <xs:annotation>
      <xs:documentation>A region in the artifact being annotated. A region is
                        the area bounded by two or more anchors. The number of anchors
required
                        to bound a region depends on the artifact being
annotated.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:attributeGroup ref="graf:graf.att.id.required.attributes"/>
      <xs:attribute name="anchors" use="required" type="xs:NMTOKENS">
        <xs:annotation>
          <xs:documentation>The anchors that bound this region. The anchors
                                attribute contains a whitespace delimited list of values
that
                                represent the anchor values. Applications are expected to
know
                                how to parse the string representation of an anchor into a
                                location in the artifact being
```

```
annotated.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
      <xs:attribute name="atype" type="xs:string">
        <xs:annotation>
          <xs:documentation>The anchor type used in the @anchors attribute. This
                            is the @name of one of the anchorTypes defined in the
header.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="link">
    <xs:annotation>
      <xs:documentation>Links a node (vertex) to a region in the artifact being
                        annotated.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:attribute name="targets" use="required" type="xs:NMTOKENS">
        <xs:annotation>
          <xs:documentation>The ID values of the regions the node is linked to. This
                            should be IDREFs, however the regions may be defined in a
                            separate file.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="a">
    <xs:annotation>
      <xs:documentation>An annotation is a labeled feature structure.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element minOccurs="0" ref="graf:fs"/>
      </xs:sequence>
      <xs:attribute name="label" use="required" type="xs:string">
        <xs:annotation>
          <xs:documentation/>
        </xs:annotation>
      </xs:attribute>
      <xs:attribute name="ref" use="required" type="xs:string">
        <xs:annotation>
          <xs:documentation/>
        </xs:annotation>
      </xs:attribute>
      <xs:attribute name="as" type="xs:string">
        <xs:annotation>
          <xs:documentation>the set this annotation belongs to, if any.</xs:documentation>
        </xs:annotation>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="as">
    <xs:annotation>
      <xs:documentation>All annotations are grouped into annotation sets.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element maxOccurs="unbounded" ref="graf:a"/>
      </xs:sequence>
      <xs:attributeGroup ref="graf:graf.att.type.attributes"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="header">
    <xs:annotation>
      <xs:documentation>groups header information found at the top of each
                        standoff annotation file. Only minimal header information is
stored in
                        the standoff file, namely a brief description of the tags
(annotations)
                        that appear and the number of times they
appear.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:choice minOccurs="0" maxOccurs="unbounded">
        <xs:element ref="graf:roots"/>
        <xs:element ref="graf:tagsDecl"/>
        <xs:element ref="graf:dependencies"/>
        <xs:element ref="graf:annotationSets"/>
```

```
      <xs:element ref="graf:layers"/>
      <xs:element ref="graf:media"/>
      <xs:element ref="graf:anchorTypes"/>
    </xs:choice>
  </xs:complexType>
</xs:element>
<xs:element name="roots">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" ref="graf:root"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="root" type="xs:token">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
</xs:element>
<xs:element name="tagsDecl">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element minOccurs="0" maxOccurs="unbounded" ref="graf:tagUsage"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="tagUsage">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
  <xs:complexType>
    <xs:attribute name="gi" type="xs:string">
      <xs:annotation>
        <xs:documentation>a generic identifier for the tag.</xs:documentation>
      </xs:annotation>
    </xs:attribute>
    <xs:attribute name="occurs" type="xs:int">
      <xs:annotation>
        <xs:documentation>the number of occurences of this tag
                                (annotation) in the file.</xs:documentation>
      </xs:annotation>
    </xs:attribute>
  </xs:complexType>
</xs:element>
<xs:element name="dependencies" type="graf:graf_dependsOn">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
</xs:element>
<xs:complexType name="graf_dependsOn">
  <xs:sequence>
    <xs:element ref="graf:dependsOn"/>
  </xs:sequence>
</xs:complexType>
<xs:element name="dependsOn">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
  <xs:complexType>
    <xs:attributeGroup ref="graf:graf.att.location.attributes"/>
  </xs:complexType>
</xs:element>
<xs:element name="media">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" ref="graf:medium"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="medium">
  <xs:annotation>
```

```
    <xs:documentation/>
  </xs:annotation>
  <xs:complexType>
    <xs:attributeGroup ref="graf:graf.att.location.attributes"/>
    <xs:attributeGroup ref="graf:graf.att.type.attributes"/>
    <xs:attributeGroup ref="graf:graf.att.name.attributes"/>
  </xs:complexType>
</xs:element>
<xs:element name="anchorTypes">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" ref="graf:anchorType"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="anchorType">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
  <xs:complexType>
    <xs:attributeGroup ref="graf:graf.att.type.attributes"/>
    <xs:attributeGroup ref="graf:graf.att.name.attributes"/>
    <xs:attribute name="media" type="xs:string">
      <xs:annotation>
        <xs:documentation/>
      </xs:annotation>
    </xs:attribute>
  </xs:complexType>
</xs:element>
<xs:element name="annotationSets">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" ref="graf:annotationSet"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="annotationSet">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element minOccurs="0" ref="graf:fs"/>
    </xs:sequence>
    <xs:attributeGroup ref="graf:graf.att.name.attributes"/>
    <xs:attributeGroup ref="graf:graf.att.type.attributes"/>
    <xs:attribute name="layers" type="xs:NMTOKENS">
      <xs:annotation>
        <xs:documentation/>
      </xs:annotation>
    </xs:attribute>
  </xs:complexType>
</xs:element>
<xs:element name="layers">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" ref="graf:layer"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="layer">
  <xs:annotation>
    <xs:documentation/>
  </xs:annotation>
  <xs:complexType>
    <xs:attributeGroup ref="graf:graf.att.name.attributes"/>
    <xs:attributeGroup ref="graf:graf.att.type.attributes"/>
    <xs:attribute name="parents" type="xs:NMTOKENS">
      <xs:annotation>
        <xs:documentation/>
```

```
        </xs:annotation>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Annex B (Informative) GrAF XML schema for the corpus header

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified"
targetNamespace="http://www.xces.org/ns/GrAF/1.0" xmlns:graf="http://www.xces.org/ns/GrAF/1.0"
xmlns:lnk="http://www.w3.org/1999/xlink">
  <xs:import namespace="http://www.w3.org/1999/xlink" schemaLocation="lnk.xsd"/>
  <xs:import namespace="http://www.w3.org/XML/1998/namespace" schemaLocation="xml.xsd"/>
  <xs:element name="cesHeader">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="graf:fileDesc"/>
        <xs:element ref="graf:encodingDesc"/>
        <xs:element ref="graf:corpusProfile"/>
      </xs:sequence>
      <xs:attribute name="creator" use="required" type="xs:NCName"/>
      <xs:attribute name="date.created" use="required" type="xs:NMTOKEN"/>
      <xs:attribute name="type" use="required" type="xs:NCName"/>
      <xs:attribute name="version" use="required" type="xs:decimal"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="fileDesc">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="graf:titleStmt"/>
        <xs:element ref="graf:editionStmt"/>
        <xs:element ref="graf:extent"/>
        <xs:element maxOccurs="unbounded" ref="graf:publicationStmt"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="titleStmt">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="graf:title"/>
        <xs:element ref="graf:version"/>
        <xs:element ref="graf:funder"/>
        <xs:element ref="graf:respStmt"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="title" type="xs:string"/>
  <xs:element name="version" type="xs:NMTOKEN"/>
  <xs:element name="funder" type="xs:string"/>
  <xs:element name="respStmt">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="graf:resp"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="resp">
    <xs:complexType mixed="true">
      <xs:attribute ref="lnk:href" use="required"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="editionStmt">
    <xs:complexType>
      <xs:attribute name="version" use="required" type="xs:NMTOKEN"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="extent">
    <xs:complexType>
      <xs:attribute name="wordCount" use="required" type="xs:integer"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="publicationStmt">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="graf:distributor"/>
        <xs:element ref="graf:pubAddress"/>
```

```
      <xs:element minOccurs="0" ref="graf:telephone"/>
      <xs:element ref="graf:fax"/>
      <xs:element ref="graf:eAddress"/>
      <xs:element ref="graf:pubDate"/>
      <xs:element ref="graf:availability"/>
      <xs:element minOccurs="0" ref="graf:idno"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="distributor" type="xs:string"/>
<xs:element name="pubAddress" type="xs:string"/>
<xs:element name="telephone" type="xs:string"/>
<xs:element name="fax" type="xs:string"/>
<xs:element name="eAddress" type="xs:string"/>
<xs:element name="pubDate">
  <xs:complexType>
    <xs:attribute name="value" use="required" type="xs:date"/>
  </xs:complexType>
</xs:element>
<xs:element name="availability">
  <xs:complexType>
    <xs:attribute name="status" use="required" type="xs:NCName"/>
    <xs:attribute ref="lnk:href" use="required"/>
  </xs:complexType>
</xs:element>
<xs:element name="idno">
  <xs:complexType>
    <xs:attribute name="type" use="required" type="xs:NCName"/>
  </xs:complexType>
</xs:element>
<xs:element name="encodingDesc">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="graf:projectDesc"/>
      <xs:element ref="graf:samplingDecl"/>
      <xs:element ref="graf:editorialDecl"/>
      <xs:element ref="graf:classDecl"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="projectDesc" type="xs:string"/>
<xs:element name="samplingDecl" type="xs:string"/>
<xs:element name="editorialDecl">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="graf:transduction"/>
      <xs:element ref="graf:correction"/>
      <xs:element ref="graf:segmentation"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="transduction" type="xs:string"/>
<xs:element name="correction" type="xs:string"/>
<xs:element name="segmentation" type="xs:string"/>
<xs:element name="classDecl">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="graf:taxonomy"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="taxonomy">
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" ref="graf:category"/>
    </xs:sequence>
    <xs:attribute name="id" use="required" type="xs:NCName"/>
  </xs:complexType>
</xs:element>
<xs:element name="corpusProfile">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="graf:fileStruct"/>
      <xs:element ref="graf:annotationSets"/>
      <xs:element ref="graf:media"/>
      <xs:element ref="graf:anchorTypes"/>
      <xs:element ref="graf:layers"/>
    </xs:sequence>
  </xs:complexType>
```

```
</xs:element>
<xs:element name="fileStruct">
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" ref="graf:fileType"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="fileType">
  <xs:complexType>
    <xs:sequence>
      <xs:element minOccurs="0" ref="graf:requires"/>
    </xs:sequence>
    <xs:attribute name="a.type" type="xs:NCName"/>
    <xs:attribute name="a.types" type="xs:NCName"/>
    <xs:attribute name="kind" type="xs:NCName"/>
    <xs:attribute name="medium" use="required" type="xs:NCName"/>
    <xs:attribute name="name" type="xs:NCName"/>
    <xs:attribute name="suffix" type="xs:NCName"/>
  </xs:complexType>
</xs:element>
<xs:element name="requires">
  <xs:complexType>
    <xs:attribute name="f.type" use="required" type="xs:NCName"/>
  </xs:complexType>
</xs:element>
<xs:element name="annotationSets">
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" ref="graf:annotationSet"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="annotationSet">
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" ref="graf:annotation"/>
    </xs:sequence>
    <xs:attribute name="type" type="xs:anyURI"/>
    <xs:attribute ref="lnk:href"/>
    <xs:attribute ref="xml:id" use="required"/>
  </xs:complexType>
</xs:element>
<xs:element name="annotation">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="graf:a.desc"/>
      <xs:element ref="graf:a.resp"/>
      <xs:element ref="graf:a.method"/>
      <xs:element maxOccurs="unbounded" ref="graf:a.doc"/>
    </xs:sequence>
    <xs:attribute name="layer" type="xs:NCName"/>
    <xs:attribute name="name" use="required" type="xs:NCName"/>
    <xs:attribute ref="lnk:href"/>
  </xs:complexType>
</xs:element>
<xs:element name="a.desc" type="xs:string"/>
<xs:element name="a.resp">
  <xs:complexType mixed="true">
    <xs:attribute ref="lnk:href" use="required"/>
  </xs:complexType>
</xs:element>
<xs:element name="a.method">
  <xs:complexType>
    <xs:attribute name="type" use="required"/>
  </xs:complexType>
</xs:element>
<xs:element name="a.doc">
  <xs:complexType>
    <xs:attribute ref="lnk:href" use="required"/>
  </xs:complexType>
</xs:element>
<xs:element name="media">
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" ref="graf:medium"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

```
  <xs:element name="medium">
    <xs:complexType>
      <xs:attribute name="encoding" use="required" type="xs:NCName"/>
      <xs:attribute name="suffix" use="required" type="xs:NMTOKEN"/>
      <xs:attribute name="type" use="required"/>
      <xs:attribute ref="xml:id" use="required"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="anchorTypes">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="graf:anchorType"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="anchorType">
    <xs:complexType>
      <xs:attribute name="default" use="required" type="xs:boolean"/>
      <xs:attribute name="medium" use="required" type="xs:NCName"/>
      <xs:attribute ref="lnk:href" use="required"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="layers">
    <xs:complexType>
      <xs:sequence>
        <xs:element maxOccurs="unbounded" ref="graf:layer"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="layer">
    <xs:complexType mixed="true">
      <xs:sequence>
        <xs:element minOccurs="0" maxOccurs="unbounded" ref="graf:l.desc"/>
      </xs:sequence>
      <xs:attribute name="name" use="required" type="xs:NCName"/>
      <xs:attribute name="ref" use="required" type="xs:NCName"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="l.desc" type="xs:string"/>
  <xs:element name="category">
    <xs:complexType>
      <xs:choice minOccurs="0" maxOccurs="unbounded">
        <xs:element ref="graf:category"/>
        <xs:element ref="graf:catDesc"/>
      </xs:choice>
      <xs:attribute name="id" use="required" type="xs:NCName"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="catDesc" type="xs:string"/>
</xs:schema>
```