Final Draft

ISO/DIN WG 6 PWI 24624

Language resource management

Transcription of spoken language

Coordinating author: Thomas Schmidt ([thomas.schmidt@ids-mannheim.de](mailto:thomas.schmidt@ids-mannheim.de))

Document history:

22-October-2012: 1st draft for ISO/DIN Meeting in Berlin, based on jTEI Paper

18-March-2013: 2nd draft for ISO/DIN Meeting in Potsdam, modifications and additions

according to discussion in Berlin and comments from the French group

20-June-2014: 3rd draft for ISO/DIN Meeting in Berlin, incorporating the result of several

discussions with the French group and comments by ISO members, filling gaps in the document, revised examples,

30-September-2014: Final draft

09-January-2015: Modifications according to suggestions from Spain (AENOR) and Slovenia (SIST), replaced <annotatedU> with <annotationGrp> to reflect the coherent use with the stand-off annotation proposal for the TEI guidelines (resp.: Laurent Romary), revised and updated examples and appendices

Table of Contents

[Foreword 4](#_Toc408578593)

[Introduction 4](#_Toc408578594)

[1 Scope 6](#_Toc408578595)

[2 Normative references 6](#_Toc408578596)

[3 Terms and definitions 7](#_Toc408578597)

[3.1 Acronyms 7](#_Toc408578598)

[4 Metadata 8](#_Toc408578599)

[4.1 Description of the electronic file (<fileDesc>) 8](#_Toc408578600)

[4.1.1 Distribution information (<publicationStmt>) 8](#_Toc408578601)

[4.1.2 Recording information (<recordingStmt>) 8](#_Toc408578602)

[4.2 Description of circumstances (<profileDesc>) 9](#_Toc408578603)

[4.2.1 Participant information (<particDesc>) 9](#_Toc408578604)

[4.2.2 Setting information (<settingDesc>) 10](#_Toc408578605)

[4.3 Description of source (<encodingDesc>) 10](#_Toc408578606)

[5 Macrostructure 12](#_Toc408578607)

[5.1 Timeline (<timeline>) 12](#_Toc408578608)

[5.2 Utterances (<u>) 12](#_Toc408578609)

[5.3 Free dependent annotations (<spanGrp>, <span>) 13](#_Toc408578610)

[5.4 Grouping of utterances and dependent annotations (<annotationGrp>) 15](#_Toc408578611)

[5.5 Independent elements outside utterances (<pause> and <incident>) 17](#_Toc408578612)

[5.6 Inline paralinguistic annotation (<shift>) 17](#_Toc408578613)

[5.7 Global divisions of a transcription (<div>) 18](#_Toc408578614)

[6 Microstructure 19](#_Toc408578615)

[6.1 Tokens (<w>) 19](#_Toc408578616)

[6.1.1 Characterisation 19](#_Toc408578617)

[6.1.2 Representation as <w> 19](#_Toc408578618)

[6.1.3 Further constraints 20](#_Toc408578619)

[6.1.4 Examples 20](#_Toc408578620)

[6.2 Pauses (<pause>) 21](#_Toc408578621)

[6.2.1 Characterisation 21](#_Toc408578622)

[6.2.2 Representation as <pause> 21](#_Toc408578623)

[6.2.3 Further constraints 21](#_Toc408578624)

[6.2.4 Examples 22](#_Toc408578625)

[6.3 Audible and visible non-speech events (<vocal>, <kinesic> and <incident>) 22](#_Toc408578626)

[6.3.1 Characterisation 22](#_Toc408578627)

[6.3.2 Representation as <vocal>, <kinesic> or <incident> 23](#_Toc408578628)

[6.3.3 Further constraints 23](#_Toc408578629)

[6.3.4 Examples 23](#_Toc408578630)

[6.4 Punctuation (<pc>) 24](#_Toc408578631)

[6.4.1 Characterisation 24](#_Toc408578632)

[6.4.2 Representation as <pc> 25](#_Toc408578633)

[6.4.3 Further constraints 25](#_Toc408578634)

[6.4.4 Examples 25](#_Toc408578635)

[6.5 Uncertainty, alternatives, incomprehensible and omitted passages (<unclear>, <choice>, <gap>) 25](#_Toc408578636)

[6.5.1 Characterisation 26](#_Toc408578637)

[6.5.2 Representation as <unclear> or <gap> 26](#_Toc408578638)

[6.5.3 Further constraints 26](#_Toc408578639)

[6.5.4 Examples 26](#_Toc408578640)

[6.6 Units above the token and below the <u> level (<seg>) 28](#_Toc408578641)

[6.6.1 Characterisation 28](#_Toc408578642)

[6.6.2 Representation as <seg> 28](#_Toc408578643)

[6.6.3 Further constraints 28](#_Toc408578644)

[6.6.4 Examples 28](#_Toc408578645)

[7 Bibliographical reference 30](#_Toc408578646)

[Annex A 32](#_Toc408578647)

[TEI-ODD specification 32](#_Toc408578648)

[Annex B 35](#_Toc408578649)

[Fully encoded example 35](#_Toc408578650)

[Annex C 39](#_Toc408578651)

[Element index 39](#_Toc408578652)

[Attribute index 39](#_Toc408578653)

# Foreword

**“The foreword of ISO documents are standard texts drafted by ISO’s Technical Management Board. ISO Central Secretariat (ISO/CS) inserts them during editintg and publishing.”**

# Introduction

This standard aims at enabling and facilitating the interchange of transcriptions of spoken language between different computational tools and environments for creating, editing, publishing and exploiting such data. Transcription of spoken language in this context means an orthography-based transcription of verbal activity as recorded in an audio or video recording of a natural interaction. The description of activity in other modalities (e.g. gestures, facial expression) may be part of a spoken language transcription, but the standard starts from the assumption that the verbal dimension is the primary focus of a spoken language transcription. Likewise, although the standard may also be relevant for transcription based on phonetic alphabets like the IPA, the assumption for this document is that orthography-based transcription is the default case.

The standard is developed in the context of the joint agreement between ISO and the TEI consortium so that its content is also distributed as part of the TEI guidelines.

The standard takes into account data models and encoding practices supported by widely used transcription software. More specifically, it builds on several interoperability studies (Schmidt et al. 2009, Schmidt et al. 2010, Parisse/Morgenstern 2010, Schmidt 2011) involving the following tools:

* ANVIL (Kipp 2014)
* CLAN (MacWhinney 2000)
* ELAN (Sloetjes 2014)
* EXMARaLDA (Schmidt/Wörner 2014)
* FOLKER (Schmidt/Schütte 2010)
* Transcriber (Barras et al. 2000)

Data encoded following this standard is expected to be compatible with the formats produced by these tools. The compatibility may extend to the formats of further labelling tools (e.g. Praat, Boersma/Weenik 1996, or Wavesurfer, http://www.speech.kth.se/wavesurfer/index2.html), but possibly on a lower level and/or with a necessity of first converting these formats to one of the above-mentioned and adding mandatory information (e.g. speaker assignment) there.

The standard also aims at being usable with widely used transcription systems (“conventions”). Compatibility in a technical sense is, however, not easily definable in this area, since, unlike the tool formats, most of these systems lack an explicit formalisation. The following selection of transcription systems was considered for this standard:

* CHAT (MacWhinney 2000)
* DT (DuBois et al. 1993)
* GAT (Selting et al. 2009)
* HIAT (Rehbein et al. 2004)

Since TEI is the reference framework for this document and metadata is not its main concern, no attempt is made here to address metadata compatibility issues beyond the TEI header. It should be noted however, that there are several TEI profiles for the CMDI framework which are related to each other and to CMDI profiles of other metadata formats (e.g. IMDI) via the ISOCAT registry (see also Broeder&Wittenburg 2002 and Broeder&van Uytvanck 2014).

This standard aims to define both a target format for legacy data conversion and a format suitable for future requirements of data processing. Individual decisions have been carefully weighed up between these two demands. At some points, certain techniques are therefore marked as preferred from a data processing point of view while an alternative technique is still allowed if the structure of legacy data makes its use unavoidable.

With regard to the other standards developed within ISO TC 37/SC 4, the present standards is intended to provide the primary layer on top of which further annotation layers may be implemented. In particular, the use of the <w> element for tokenizing a transcription is conformable to the TEI based representation of tokens ISO 24611 (MAF).

This standard also aligns with the mechanism proposed in the TEI guidelines to embed stand-off annotations within a TEI document. This mechanism contains in particular the provision of a generic element (<annotationGrp>) to group together annotations related to the same linguistic segment, which fits the needs of the present standard in the case of annotated <u> elements.

# 1 Scope

This document specifies rules for representing transcriptions of audio- or video-recorded spoken interactions in XML documents based on the guidelines of the Text Encoding Initiative. As a secondary objective, the document aims at relating transcribed data with standards for annotated corpora.

It is applicable to transcription data for studies in sociolinguistics, conversation analysis, dialectology, corpus linguistics, corpus lexicography, qualitative social studies and other transcription data of recorded spoken language.

It is not applicable to other forms of transcription, most importantly transcriptions of hand-written manuscripts.

# 2 Normative references

This standard and the TEI guidelines are based on SGML/XML (ISO 8879) and the Universal Character Set (Unicode) (ISO 10646).

The TEI version used is TEI P5, version 2.6.0 of 2014-01-20.

As a TEI application, this document refers to the following standards:

* Data elements and interchange formats – Information interchange – Representation of dates and times (ISO 8601)
* Language codes (ISO 639)
* Information technology — Codes for the representation of human sexes (ISO 5218)

In addition, the **<media>** element has an attribute **@mimeType** which makes use of Multipurpose Internet Mail Extensions (MIME) as specified in IETF RFC #2046 (<http://www.ietf.org/rfc/rfc2046.txt>) and an attribute **@url** which makes use of Uniform Resource Locators (URLs) as specified in in IETF RFC #1738 (<http://www.ietf.org/rfc/rfc1738.txt>).

On the level of tokens, this standard is compatible with ISO 24611:2012, “Language resource management — Morpho-syntactic annotation framework (MAF)”.

The use of the <annotationGrp> element is compatible with … Laurent?

# 3 Terms and definitions

## 3.1 Acronyms

**CHAT**  Codes for the Human Analysis of Transcripts (guideline)

**CLAN**  Computerized Language Analyses (tool)

**CMDI**  Component Metadata Initiative (metadata framework)

**DT**  Discourse Transcription (guideline)

**ELAN** EUDICO (=European Distributed Corpora) Linguistic Annotator (tool)

**EXMARaLDA** Extensible Markup Language for Discourse Annotation (toolset)

**FOLKER** FOLK (=Forschungs- und Lehrkorpus Gesprochenes Deutsch, Research and Teaching Corpus of Spoken German) Editor (tool)

**GAT** Gesprächsanalytisches Transkriptionssystem (Conversation Analytic Transcription System) (guideline)

**HIAT** Halbinterpretative Arbeitstranskriptionen (Semi-Interpretative Working Transcriptions) (guideline)

**IPA** International Phonetic Alphabet

**MAF** Morphosyntactic Annotation Framework

**POS** Part of speech

**TEI** Text Encoding Initiative

# 4 Metadata

The TEI Guidelines formulate extensive suggestions for encoding metadata inside different subsections of the **<teiHeader>** element. The following section addresses only those pieces of metadata which are either crucial for ensuring the interpretability and exchangeability of spoken language transcriptions in general or which can be expected to be relevant in a large majority of cases. This does not preclude the possibility of or necessity for encoding further metadata inside the **<teiHeader>** element.

## 4.1 Description of the electronic file (<fileDesc>)

### 4.1.1 Distribution information (<publicationStmt>)

The **<publicationStmt>** element inside the **<fileDesc>** section of the **<teiHeader>** should be used to record information about access rights and contact information for the transcription in question.

|  |
| --- |
| <publicationStmt>  <authority>Hamburger Zentrum für Sprachkorpora</authority>  <availability>  <licence target="http://www.corpora.uni-hamburg.de/licence.html"/>  <p>Available free for research and teaching purposes.  No redistributing allowed. </p>  </availability>  <distributor>Hamburger Zentrum für Sprachkorpora</distributor>  <address>  <street>Max Brauer-Allee 60</street>  <postCode>22765</postCode>  <placeName>Hamburg</placeName>  <country>Germany</country>  </address>  </publicationStmt> |

### 4.1.2 Recording information (<recordingStmt>)

The **<recordingStmt>** element inside the **<fileDesc>** section of the **<teiHeader>** should be used to record information about the transcribed recording(s). A **<media>** element inside a **<recording>** element should be used to refer to the corresponding digital file via a **@url** attribute. A **@type** attribute on **<recording>** should be used to indicate the media type of the recording. *audio* and *video* are the permissible values for that attribute. The actual digital file type should be encoded as a **@mimeType** attribute on the **<media>** element. Where two or more files are derived from the same master recording, these should be represented as different **<media>** elements inside the same **<recording>** element, rather than as different **<recording>** elements.

|  |
| --- |
| <recordingStmt>  <recording type="video">  <!-- element from TEI P5, but not allowed there as a child of recording -->  <media mimeType="video/mpeg" url="Beckhams.mpg"/>  <media mimeType="audio/wav" url="Beckhams.wav"/>  <broadcast>  <ab>Parkinson Talkshow on BBC, broadcast on 02 November 2007</ab>  </broadcast>  <!-- information about the equipment used for creating the recording -->  <!-- where recordings are made by the researcher, this would be -->  <!-- place to specify the recording equipment (e.g. Camcorder) -->  <equipment>  <ab>Video excerpt downloaded from YouTube with aTube-Catcher, converted  into MPG format with Adobe Premiere</ab>  <ab>Audio extracted from video with Audacity 1.3 beta</ab>  </equipment>  </recording>  </recordingStmt> |

## 4.2 Description of circumstances (<profileDesc>)

### 4.2.1 Participant information (<particDesc>)

The participants of the transcribed interaction should be described in **<person>** elements inside the **<particDesc>** section of a **<profileDesc>** element. Using an **@n** attribute on the **<person>** element to define an abbreviated code for the respective participant is mandatory since it can be crucial for many processing purposes. **<u>** elements inside the body of the transcription refer to the **@xml:id** attribute of a **<person>** element which must therefore always be provided.

In order to provide additional metadata about participants, the content model of **<person>** can be fully exploited, for example to record a person’s age, birth date, language knowledge, etc.

|  |
| --- |
| <particDesc>  <person xml:id="SPK0" sex="1" n="DS">  <persName>  <forename>Daniel</forename>  <surname>Steward</surname>  </persName>  <age value="34"/>  <birth when="1960-12-10"/>  <langKnowledge>  <langKnown tag="en-GB" level="H">British English</langKnown>  <langKnown tag="fr" level="M">French</langKnown>  </langKnowledge>  <!-- possibly further descriptive elements -->  </person>  <person xml:id="SPK1" sex="2" n="FB">  <persName>  <forename>Fiona</forename>  <surname>Baker</surname>  </persName>  <!-- possibly further descriptive elements -->  </person>  </particDesc> |

### 4.2.2 Setting information (<settingDesc>)

The **<settingDesc>** element should be used to provide information about the place and time, spatial organization, artifacts, etc. of the transcribed interaction.

|  |
| --- |
| <settingDesc>  <place>  <placeName>BBC studio London</placeName>  </place>  <ab>  <date when="2009-02">early February 2009</date>  </ab>  <setting>  <activity>Talkshow host Michael Parkinson interviewing David and Victoria  Beckham about their relationship</activity>  </setting>  <!-- possibly further descriptive elements -->  </settingDesc> |

## 4.3 Description of source (<encodingDesc>)

The **<encodingDesc>** element should be used to record information about the source from which the TEI document was derived. This includes information about the tool which created the transcription inside an **<appInfo>** element and information about the convention which was used in transcribing the data inside a **<transcriptionDesc>** element. **@ident** and **@version** attributes should be used on these elements to provide a machine-readable way of accessing this information.

|  |
| --- |
| <encodingDesc>  <appInfo>  <!-- information about the application with which -->  <!-- the transcription was created -->  <application ident="EXMARaLDA" version="1.5.1">  <label>EXMARaLDA Partitur-Editor</label>  <desc>Transcription Tool providing a TEI Export</desc>  </application>  </appInfo>  <!-- information about the transcription convention used -->  <transcriptionDesc ident="HIAT" version="2004">  <desc>Orthographic transcription according to HIAT</desc>  </transcriptionDesc>  </encodingDesc> |

# 5 Macrostructure

## 5.1 Timeline (<timeline>)

**<when>** elements inside a **<timeline>** element should be used to define points in the recording which are then referred to by **@start**, **@end** and **@synch** attributes of other elements (most importantly **<anchor>** elements) of the transcription to represent its temporal structure. It is therefore obligatory to provide an **@xml:id** attribute for each **<when>** element. **<when>** elements must be in the same order as the timepoints they refer to. Specifying an offset into the recording via an **@interval** attribute is optional, but very useful for many processing purposes. Absolute time values in the **@interval** attribute should be given in seconds with as many decimal digits as necessary.

|  |
| --- |
| <timeline unit="s" origin="#T0">  <when xml:id="T0" absolute="00:00:00.0"/>  <when xml:id="T1" interval="2.13" since="#T0"/>  <when xml:id="T2" interval="3.74" since="#T0"/>  <when xml:id="T3" interval="4.71" since="#T0"/>  <when xml:id="T4" interval="unknown" since="#T0"/>  <when xml:id="T5" interval="8.53" since="#T0"/>  <when xml:id="T6" interval="11.36" since="#T0"/>  <when xml:id="T7" interval="13.91" since="#T0"/>  <when xml:id="T8" interval="15.47" since="#T0"/>  <!-- [...] more when elements -->  </timeline> |

## 5.2 Utterances (<u>)

The **<u>** element is the fundamental unit of organization for a transcription, roughly comparable to a paragraph (**<p>** element) in a written document. It corresponds to a contiguous stretch of speech of a single speaker. A more exact definition and delimitation of a **<u>** is not in the scope of this document. The TEI definition characterising a **<u>** as “often preceded by a silence or a change of speaker” should be viewed as a suggestion only. So it is permissible to use a more refined definition for a <**u>.** This more refined definition can be described in the header in a **<transcriptionDesc>** element inside an **<encodingDesc>** element.

If they are not wrapped inside an **<annotationGrp>** element (see section 5.4), **<u>** elements must be assigned to a single speaker by providing a value for the **@who** attribute which points to the **@xml:id** of a **<person>** element defined in the header. If the speaker cannot be identified, the **@who** attribute may also be omitted. An **@xml:id** attribute can optionally serve to make the **<u>** element addressable for standoff-annotation, for instance via **<span>** elements (see below).

If they are not wrapped inside an **<annotationGrp>** element (see section 5.4), **<u>** elements must be assigned to the timeline by providing values for the **@start** and **@end** attributes pointing to the the **@xml:id** of a **<when>** element defined in the timeline. Further temporal structure can be recorded by inserting **<anchor>** elements at appropriate places inside the content of a **<u>** element.

The preferred mechanism for representing overlap is to encode it implicitly through the appropriate use of **@start** and **@end** attributes and **<anchor>** elements. Other TEI mechanisms, such as a **@trans=’overlap’** attribute for the **<u>** element, are allowed but not recommended because they cannot be processed in an appropriate manner by many of the widely used annotation tools.

|  |
| --- |
| <!-- u with start and end attributes only (minimal temporal structure) -->  <u who="#SPK1" start="#T0" end="#T1" xml:id="u2">Good morning! </u>  <!-- u with embedded anchor elements (additional temporal structure) -->  <u who="#SPK0" start="#T1" end="#T4">  Okay. <anchor synch="#T2"/>Très bien, <anchor synch="#T3"/>très bien.  </u>  <!-- two <u>s with partial overlap -->  <u who="#SPK0" start="#T0" end="#T2">Do not <anchor synch="#T1"/>interrupt me!</u>  <u who="#SPK1" start="#T1" end="#T3">Sorry, <anchor synch="#T2"/>mate!</u> |

In the simplest case, **<u>** elements contain character data, possibly interspersed with **<anchor>** elements (see examples above). Further structuring of the content of a **<u>** element (e.g. markup of tokens, pauses, etc.) may be done via the mechanisms described in section 6.

The assumed default case is that **<u>** contains an orthographic transcription in a broad sense, including orthography-based mechanisms for approaching the actual phonetic realisations, such as “eye dialect”, “literary transcription” or “modified orthography”. If this is the case, no further specification in the form of a **@type** attribute on **<u>** is necessary. If, however, **<u>** contains a phonemic or phonetic transcription or is based on some other systematics, this should be indicated via a **@type** attribute with an appropriate value.

|  |
| --- |
| <!-- u with phonetic transcription in IPA -->  <u who="#SPK1" start="#T0" end="#T1" type="phonetic">ɡʊd ˈmɔːnɪŋ</u> |

If several types of transcription exist side-by-side (e.g. an orthographic and a phonetic transcription), one level should be singled out as the primary transcription layer. Only this layer should be represented inside **<u>** elements, the other one being represented in appropriate **<span>** elements (see section 5.3).

## 5.3 Free dependent annotations (<spanGrp>, <span>)

Whereas **<u>** (typically, but not necessarily) contains the basic orthographic transcription, **<span>** elements should be used to represent additional annotations (e.g. POS tagging, prosodic annotation, translation) on that basic transcription. Annotations of the same type should be grouped in a **<spanGrp>** element with a **@type** attribute specifying the annotation level.

The reference of the annotation in question must be specified using **@to** and **@from** attributes in one of the following ways:

* the values of **@to** and **@from** can point to the **@xml:id** attributes of other elements (e.g. a **<u>**, a **<w>** or a **<seg>**) of the transcription
* the values of **@to** and **@from** can point to the **@xml:id** attributes of **<when>** elements from the timeline

If the latter mechanism is used, **<spanGrp>** elements must be grouped with the **<u>** element they refer to by using an **<annotationGrp>** element (see section 5.4).

On the level of tokens, annotation via **<span>** elements pointing to **<w>** elements is conformable to the annotation mechanism described in ISO 24611 (MAF).

Annotations of single tokens, such as lemmatisation, POS tagging, etc., may alternatively be realised as appropriate attributes on **<w>** elements if no structural conflicts between the two levels exist (see 7.1.2.).

For annotations with an internal structure, nesting **<span>** elements can be used. In that way, 1:n relations between tokens and annotations, as well as hierarchically organized annotations can be expressed.

The use of further annotation techniques (e.g. via feature structures) is not precluded, but not in the scope of this document.

|  |
| --- |
| <!-- annotations from a sup (=suprasegmentals) tier -->  <!-- using a reference to the timeline -->  <spanGrp type="sup">  <span from="#T2" to="#T4">faster</span>  </spanGrp>  <!-- annotations from an en (=English translation) tier -->  <!-- using a reference to the timeline -->  <spanGrp type="en">  <span from="#T1" to="#T2">Okay. </span>  <span from="#T2" to="#T4">Very good, very good.</span>  </spanGrp>  <!-- part-of-speech annotations -->  <!-- using a reference to ids of <w> elements -->  <spanGrp type="pos">  <span from="#w148" to="#w148">PersPron</span>  </spanGrp>  <!— 1:n relation between tokens and annotations -->  <u><w xml:id="w1">I</w><w xml:id="w2">dunno</w></u>  <spanGrp type="lemma">  <span from="#w1" to="#w1">I</span>  <span from="#w2" to="#w2">  <span>do</span>  <span>not</span>  <span>know</span>  </span>  </spanGrp>  <!— hierarchically organised annotation -->  <u>  <w xml:id="w3">John</w><w xml:id="w4">loves</w><w xml:id="w5">Mary</w>  </u>  <spanGrp type="phraseStructure">  <span from="#w3" to="#w5">  <span>S</span>  <span from="#w3" to="#w3">  <span>NP</span>  <span from="#w3" to="#w3">N</span>  </span>  <span from="#w4" to="#w5">  <span>VP</span>  <span from="#w4" to="#w4">V</span>  <span from="#w5" to="#w5">  <span>NP</span>  <span from="#w5" to="#w5">N</span>  </span>  </span>  </span>  </spanGrp> |

## 5.4 Grouping of utterances and dependent annotations (<annotationGrp>)

**<u>** elements and the annotations referring to them can be grouped under an **<annotationGrp>** element. This has the advantage of creating “local” annotated environments each (succession) of which can be treated as an independent transcription in its own right (“tesselation” of the transcription document). **<spanGrp>** elements in which spans point to the timeline rather than directly to other elements of the transcription must be grouped with the **<u>** element they refer to, because otherwise ambiguities with respect to their scope may arise in cases of overlapping speech.

Although the use of **<annotationGrp>** is optional, it is not allowed to mix **<annotationGrp>** and **<u>** elements on the top level - i.e. as soon as one **<annotationGrp>** element is used, all **<u>** elements have to be wrapped inside an **<annotationGrp>** element.

**<annotationGrp>** elements must not contain more than one **<u>** element. There may be cases, however, where it makes sense to use an **<annotationGrp>** as a container only for the description of a non-verbal action of a participant, i.e. without a subordinate **<u>** element.

If **<annotationGrp>** is used as the top level element, speaker assignment through the **@who** attribute should be made on this level instead of on the embedded **<u>** element. The same holds for **@start** and **@end** attributes pointing to the timeline. An **@xml:id** attribute can be used to make the **<annotationGrp>** addressable for stand-off annotations.

The **<annotationGrp>** element can also be used as a stand-off annotation component within the <annotations> element of <stdf>, as specified in the TEI guidelines. In such a case, **<annotationGrp>** points to the corresponding <u> element by means of a @corresp attribute (@target?).

A paragraph here about the general possibility of using <annotationGrp> outside the <text> element in a “real stand-off” <stdf> element? This also as a possibility to encode annotations that go across <u> boundaries. Laurent?

|  |
| --- |
| <!-- an utterance grouped with corresponding annotations -->  <annotationGrp who="#SPK0" start="#T0" end="#T1">  <!-- the transcribed text from the primary tier -->  <u>  <!-- [...] (see above) -->  </u>  <!-- additional annotations from a sup (=suprasegmentals) tier -->  <spanGrp type="sup">  <!-- [...] (see above) -->  </spanGrp>  <!-- additional annotations from an translation tier -->  <!-- with an xml:lang attribute capturing the language of the translation -->  <spanGrp type="translation" xml:lang="en">  <!-- [...] (see above) -->  </spanGrp>  </annotationGrp>  <!-- an annotationGrp without subordinate <u> element -->  <annotationGrp who="#SPK0" start="#T0" end="#T1">  <vocal>  <desc>laughter</desc>  </vocal>  </annotationGrp> |

## 5.5 Independent elements outside utterances (<pause> and <incident>)

**<pause>** and **<incident>** elements should be used to represent pauses and non-verbal phenomena which cannot be attributed to a speaker. In the document, these elements appear on the same hierarchical level as **<annotationGrp>** (or, as the case may be, **<u>**) elements. In order to fit them into the temporal structure they must have **@start** and **@end** attributes pointing to the timeline.

|  |
| --- |
| <annotationGrp who="#SPK0" start="#T0" end="#T1">  <!-- [...] u and spanGrp elements, see above -->  </annotationGrp>  <!-- an incident not attributable to a speaker -->  <incident start="#T1" end="#T2">  <desc>roar of thunder outside</desc>  </incident>  <!-- a pause not attributable to a speaker -->  <pause dur="PT0.61S" start="#T2" end="#T3"/>  <annotationGrp who="#SPK1" start="#T3" end="#T4">  <!-- [...] u and spanGrp elements, see above -->  </annotationGrp> |

## 5.6 Inline paralinguistic annotation (<shift>)

The TEI guidelines offer the **<shift>** element to “[mark] the point at which some paralinguistic feature of a series of utterances by any one speaker changes”. If used for that purpose, the element must be further specified by the attributes **@feature** (legal values: *tempo* for speed of utterance, *loud* for loudness, *pitch* for pitch range, *tension* for tension or stress pattern, *rhythm* for rhythmic qualities and *voice* for voice quality) and **@new** to provide the new value taken by the feature at this point. In addition, a **@synch** attribute must be provided to assign the element a position in the timeline.

**<shift>** is a milestone element. As such, it brings with it certain problems with automatic checking and processing of the document structure. Since the description of paralinguistic features can also be viewed as annotations of transcribed material, expressing the same content in a **<span>** element (see section 5.3) is the preferable alternative.

|  |
| --- |
| <!-- a change of tempo encoded as a <shift> milestone -->  <u start="#T1" end="#T4" who="#SPK1">  And he was <shift feature="tempo" new="faster" synch="#T2"/>up and away  <shift feature="tempo" new="normal" synch="#T4"/>  </u>  <!-- the same phenomenon encoded as an annotation in a <span> -->  <annotationGrp start="#T1" end="#T4" who="#SPK1">  <u>  And he was <anchor synch="#T2"/>up and away  </u>  <spanGrp type="sup">  <span from="#T2" to="#T4">faster</span>  </spanGrp>  </annotationGrp> |

## 5.7 Global divisions of a transcription (<div>)

For a division of a transcription into larger sections (above the level of **<u>** or **<annotationGrp>** elements), e.g. for different phases of an interaction, the **<div>** element with an appropriate **@type** (if needed, a **@subtype** in addition) attribute can be used. This is optional and need not be applied exhaustively.

|  |
| --- |
| <!-- initial section of the interaction -->  <div type="greeting">  <annotationGrp who="#SPK0" start="#T0" end="#T1">  <!-- [...] u and spanGrp elements, see above -->  </annotationGrp>  <annotationGrp who="#SPK1" start="#T1" end="#T2">  <!-- [...] u and spanGrp elements, see above -->  </annotationGrp>  </div>  <!-- main part, not embedded in a div element -->  <annotationGrp who="#SPK0" start="#T2" end="#T3">  <!-- [...] u and spanGrp elements, see above -->  </annotationGrp>  <!-- [...] -->  <!-- final section of the interaction -->  <div type="farewell">  <annotationGrp who="#SPK1" start="#T112" end="#T113">  <!-- [...] u and spanGrp elements, see above -->  </annotationGrp>  <annotationGrp who="#SPK0" start="#T113" end="#T114">  <!-- [...] u and spanGrp elements, see above -->  </annotationGrp>  </div> |

# 6 Microstructure

## 6.1 Tokens (<w>)

### 6.1.1 Characterisation

Most transcription conventions do not provide an exact and comprehensive definition of the unit *word*. Rather, they depart from the word definition of standard written orthography and supplement this with rules for a selected number of special cases (e.g. words specific to spoken language like ‘ehm’, abbreviations, spellings, etc.). A more precise definition should and need not be attempted in this document - the decision of what is to be treated (i.e. marked up) as a word can be left to the individual transcription system. The definition of **<w>** elements in spoken language transcription can thus be viewed as analogous to the definition of a token in the Morpho-Syntactic Annotation Framework (MAF) where “the description of the orthographic, morphological, phonological and lexical structures that may define a token is not covered by [the] standard. ” Henceforth, we will call the entity marked-up as a **<w>** element a token in order to avoid confusion with (orthographic) words in a less formal sense.

### 6.1.2 Representation as <w>

Tokens (as defined by the transcription system used) should be encoded as **<w>** elements underneath a **<u>** element. In order to make tokens referenceable in annotations, the use of an **@xml:id** attribute is recommended.

A **@type** attribute can be used to represent special features of a token, especially when the corresponding distinction is an integral part of the transcription system. For instance, the following distinctions made by several widely used transcription systems can be encoded in a **@type** attribute of a **<w>** element:

* **@type=’assimilated’** on the later word for assimilated words
* **@type=’truncated’** for truncated words
* **@type=’repetition’** for repeated words

An **@ana** attribute can serve as a place to encode the part-of-speech of the token. Similarly, a **@lemma** or **@lemmaRef** attribute can be used to associate the token with a lemma, such as an uninflected dictionary entry form.

Since information encoded in **@type**, **@ana**, **@lemma**, **@lemmaRef** attributes constitutes an annotation on the token, this kind of information can alternatively be recorded as a (free) annotation in a **<span>** element (see section 5.3). This is especially advisable if there is not a 1:1 relationship between **<w>** elements and annotations on the lemma or POS level.

Beneath the level of tokens, many transcription conventions contain instructions for marking a given syllable as accentuated/stressed or a given sound as lengthened. To delimit such units below the token level, a **<seg>** element can be used and either be characterised as an accentuated syllable or lengthened sound by an appropriate **@type** attribute or, again, by referencing the **<seg>** element from a **<span>** via its **@xml:id** attribute. If a transcription system provides a systematic and exhaustive subdivision of tokens into morphemes, the **<m>** element can be used to represent this subdivision.

### 6.1.3 Further constraints

Since overlaps starting or ending inside a token occur, **<w>** must allow **<anchor>** as a child. Pauses inside tokens can occur and should be encoded as **<pause>** elements as described in section 6.2.

### 6.1.4 Examples

|  |
| --- |
| <!-- an utterance divided into tokens -->  <u who="#SPK0" start="#T0" end="#T2">  <w xml:id="w148">I</w>  <w xml:id="w149">am</w>  <w xml:id="w150">very</w>  <w xml:id="w151">much</w>  <w xml:id="w152">aware</w>  <w xml:id="w153">of</w>  <w xml:id="w154">that</w>  </u>  <!-- token marked as assimilated via a type attribute -->  <u who="#SPK0" start="#T0" end="#T1">  <w xml:id="w1">what</w>  <w xml:id="w2" type="assimilated">cha</w>  <w xml:id="w3">got</w>  <w xml:id="w4">cookin</w>  </u>  <!-- POS and lemma information encoded as attributes on the token -->  <u who="#SPK0" start="#T0" end="#T2">  <w xml:id="w148" lemma="I" ana="PRO">I</w>  <w xml:id="w149" lemma="be" ana="V">am</w>  <w xml:id="w150" lemma="very" ana="ADV">very</w>  <w xml:id="w151" lemma="much" ana="ADV">much</w>  <w xml:id="w152" lemma="aware" ana="ADJ">aware</w>  <w xml:id="w153" lemma="of" ana="PREP">of</w>  <w xml:id="w154" lemma="that" ana="PRO">that</w>  </u>  <!-- a token with an accentuated syllable -->  <!-- the accentuation being represented in a separate span element -->  <annotationGrp who="#SPK0" start="#T0" end="#T2">  <u>  <!-- [...] -->  <w xml:id="w152"><seg xml:id="seg152a"/>awe</seg>some</w>  <!-- [...] -->  </u>  <!-- [...] -->  <spanGrp type="prosody">  <span from="#seg152a" to="#seg152a">accentuated</span>  </spanGrp>  </annotationGrp>  <!-- the same phenomenon encoded inline -->  <w xml:id="w152"><seg type="accentuated"/>awe</seg>some</w>  <!-- a token with a short pause inside -->  <w xml:id="w152">abso<pause type="short"/>lutely</w>  <!-- a token with a time anchor inside -->  <w xml:id="w152">a<anchor synch="#T3"/>ware</w> |

## 6.2 Pauses (<pause>)

### 6.2.1 Characterisation

Most transcription systems distinguish measured pauses and typed pauses, the latter being typically divided into a small number of types based on perceived length, such as ‘micro’, ‘short’, ‘medium’ and ‘long’. Pauses can occur outside speaker’s utterances (see section 5.5) and between or inside tokens attributed to a **<u>** element. Whether or not, and how, a pause is attributed to a speaker is a decision of the transcription system.

### 6.2.2 Representation as <pause>

All pauses should be represented as **<pause>** elements. For measured pauses, the length should be provided in a **@dur** attribute. For typed pauses, the type should be provided in a **@type** attribute. If neither measured length nor a typification are provided, the **<pause>** element can also be used without attributes. Since notation of pauses in legacy documents varies greatly, it may be advisable to keep the original notation form. A **@rend** attribute can be used for that purpose. As described above, pauses outside **<u>** elements need a **@start** and an **@end** attribute referring to the timeline. For pauses inside **<u>** elements, timing information can, but need not, be provided via preceding and/or following **<anchor>** elements.

### 6.2.3 Further constraints

Since the measured duration of a pause is also temporal information, contradictions may arise between the value of the **@dur** attribute and information encoded in timeline references, for instance when a pause is longer than the utterance in which it is contained. Such inconsistencies cannot be detected by document grammars.

### 6.2.4 Examples

|  |
| --- |
| <!-- measured pause -->  <pause dur="PT1.2S"/>  <!-- typed pause -->  <pause type="micro"/>  <!-- typed pause with original form in a rend attribute-->  <pause type="micro" rend="(.)"/>  <!-- measured pause outside <u>, with its own start and end attributes -->  <pause dur="PT0.61S" start="#T10" end="#T11"/> |

## 6.3 Audible and visible non-speech events (<vocal>, <kinesic> and <incident>)

### 6.3.1 Characterisation

Non-speech events comprise a very varied set of phenomena, ranging from productions with an obvious communicative function (such as audible laughter or a visible shake of the head) over (assumedly) secondary modes of communication (such as gestures or facial expressions) to events (such as “telephone rings”) and activities (such as “rummages in pocket”) which are not directly communicative but may still be crucial to an understanding of a transcribed interaction. Different transcription systems have different rules for classifying and describing such events, and it is not easy to define the common ground between them. However, a few essential distinctions seem to be relevant for all systems:

* audible (“cough”) vs. visible (“nod”) events
* events alternative (laughter at the end of an utterance) vs. events simultaneous (words uttered laughing) to speech
* events which can (“cough”, “nod”, “laughter”) vs. events which cannot (“telephone rings”, “microphone topples over”) be attributed to a speaker

Most systems will at least contain instructions for audible events which are alternative to speech and which can be attributed to a speaker. Among such phenomena the most commonly described in transcription conventions are breathing and laughing (both of which often obtain a specialized transcription symbol of their own), throat clearing, smacking noises, yawns, coughs and sneezes. If transcriptions are based on video rather than audio, conventionalized gestures such as a nod or shake of the head, a knitting of the brows, or a “thumbs up” are usually the first to be added to the repertoire of non-speech events considered in the conventions.

Since a true multimodal annotation (i.e. a systematic and exhaustive description of non-verbal behaviour) is outside the scope of this document, we will limit ourselves to instructions on how to encode these basic types of non-speech events.

### 6.3.2 Representation as <vocal>, <kinesic> or <incident>

The TEI guidelines offer three different elements for describing non-speech events (see chapter 8):

* **<vocal>** for vocalized but non-lexical phenomena such as coughs
* **<kinesic>** for kinesic (non-verbal, non-lexical) communicative phenomena such as gestures
* **<incident>** for entirely non-linguistic incidents occurring during and possibly influencing the course of speech

Most of the non-speech phenomena described in “classical” (i.e. audio-based) transcription systems will fall into the **<vocal>** class, the (video-based) description of conventionalized gestures will usually be an instance of **<kinesic>**, so that **<incident>** can be reserved for making notes of (audible or visible) not directly communicative events that may be relevant to the interaction.

**<vocal>** and **<kinesic>** elements that are alternative to speech can be embedded inside **<u>** elements if the transcription system allows or prescribes this. The speaker assignment is then inherited from the superordinate **<u>** element, no independent assignment to the timeline is required.

If they are (partly) simultaneous to an utterance by the same speaker, they can be grouped within the same **<annotationGrp>**, but outside the **<u>** element. In this case, **@start** and **@end** attributes have to be provided.

If they occur in isolation (i.e. without preceding or following lexical material), or are viewed as occurring outside the boundaries of utterances, they will have to be represented on the same hierarchical level as **<u>** or **<annotationGrp>** elements. In this case, a speaker assignment has to be encoded explicitly via a **@who** attribute, and a reference to the timeline via **@start** and **@end** attributes is mandatory.

### 6.3.3 Further constraints

### 6.3.4 Examples

|  |
| --- |
| <!-- coughing encoded as vocal element between tokens and anchors of a u -->  <u who="#SPK0" start="#T4" end="#T6">  <anchor synch="#T4"/>  <w>dépend</w>  **<vocal>**  **<desc>cough</desc>**  **</vocal>**  <anchor synch="#T5"/>  <w>un</w>  <w>peu</w>  <anchor synch="#T6"/>  </u>  <!-- simultaneous laughter by the same speaker -->  <!-- encoded as vocal element within the same annotationGrp -->  <!-- with start and end points -->  <annotationGrp who="#SPK0" start="#T4" end="#T6">  <u>  <anchor synch="#T4"/>  <w>dépend</w>  <anchor synch="#T5"/>  <w>un</w>  <w>peu</w>  <anchor synch="#T6"/>  </u>  **<vocal start="#T4" end="#T6">**  **<desc>laughing</desc>**  **</vocal>**  </annotationGrp>  <!-- (backchannel) nodding as kinesic element on the level of annotationGrp -->  <!-- with speaker assignment and start and end points -->  <annotationGrp who="#SPK0" start="#T6" end="#T9">  <!-- [...] -->  </annotationGrp>  <kinesic who="#SPK1" start="#T7" end="#T8">  <desc>nods</desc>  </kinesic> |

## 6.4 Punctuation (<pc>)

### 6.4.1 Characterisation

Since spoken utterances rarely follow the grammar of the written standard, few transcription systems employ punctuation according to standard orthography rules, e.g. a period to mark the end of a grammatical sentence or a comma to introduce a subordinate clause in German. More frequently, the semantics of punctuation symbols are redefined to match salient characteristics of spoken language. One common system is based on prosody and uses punctuation symbols to delimit intonation phrases and to characterise their final tone movement. In the German GAT system, for instance, a period marks the end of an intonation phrase with a low falling tone movement, the question mark the end of a phrase with a high rising tone movement, etc. Other uses of punctuation symbols include the marking of repair sequences (e.g. a forward slash is used in HIAT for that purpose) containing truncated words (e.g. a hyphen) and similar phenomena. Ideally, such punctuation symbols should be regarded as visual representations of annotations and should accordingly be mapped to appropriate markup such as a **@type** attribute on a **<w>** element (for truncation represented by a hyphen, see section 6.1.2) or a **@type** attribute on a **<seg>** element (for tone movements, see section 6.6). However, due to ambiguous or unclear rules in legacy systems, this may not always be feasible. If this is the case (or if the punctuation does indeed follow standard orthography rules), the punctuation symbol should be represented as such at the position at which it occurs inside a **<u>** element.

### 6.4.2 Representation as <pc>

The **<pc>** element should be used to represent punctuation characters which cannot be mapped to an annotation element or attribute. The **@type** and **@unit** attributes can be used to provide additional information about its function.

### 6.4.3 Further constraints

In contrast to other elements, a punctuation symbol does usually not correspond directly to some event occurring in time. It is therefore not possible to place it on the timeline via a **@start** and **@end** attribute or via preceding or following **<anchor>** elements.

### 6.4.4 Examples

|  |
| --- |
| <!-- punctuation represented as pc elements -->  <u who="#SPK0" start="#T4" end="#T6">  <w xml:id="w330">No</w>  <pc>,</pc>  <w xml:id="w331">I</w>  <w xml:id="w332">mean</w>  <w xml:id="w333">I</w>  <w xml:id="w334">knew</w>  <pc type="declarative">.</pc>  </u> |

## 6.5 Uncertainty, alternatives, incomprehensible and omitted passages (<unclear>, <choice>, <gap>)

### 6.5.1 Characterisation

Most transcription systems have mechanisms to mark uncertainty in transcription, i.e. parts where the transcriber is not sure of what he heard, and to identify incomprehensible passages, i.e. parts which the transcriber did not understand at all. Related to the latter are parts which may be understandable, but which the transcriber consciously decided not to transcribe.

Uncertain passages will still contain transcribed words, but it is important to be able to indicate their uncertain status. Several transcription systems allow the transcriber to offer one or more alternative transcriptions for these cases.

### 6.5.2 Representation as <unclear> or <gap>

An **<unclear>** element can be used to indicate uncertainty of a transcribed sequence of words. The **@reason** attribute can be used to provide information about the cause of the uncertainty. If more than one transcription for the uncertain passage is plausible, all possible alternatives should be represented inside a **<choice>** element subordinate to the **<unclear>** element. If there is a choice only between different single words, these words can simply be enumerated. If the choice is about sequences of words, each sequence needs to be grouped in a **<span>** element.

Completely incomprehensible passages should be represented by a **<gap>** element. The **@reason** attribute should then be attributed the value *incomprehensible*. A **@dur** attribute may be used to indicate the temporal duration of the passage. Alternatively or in addition, attributes from the **att.dimensions** class (e.g. **@unit** + **@quantity** or **@extent**) can also be used to give information about the extent of the gap.

Passages which were left untranscribed for some other reason should also be represented in a **<gap>** element with appropriate **@reason** and/or **@dur** attributes.

### 6.5.3 Further constraints

**<gap>** elements may occur inside **<u>** elements if the incomprehensible or untranscribed passage is short and clearly forms part of an utterance of which other parts have been transcribed, or it may occur on the same level as **<u>** or **<annotationGrp>** elements if the omission is of a more global nature. In the latter case, **@start** and **@end** attributes pointing to the timeline must be provided.

### 6.5.4 Examples

|  |
| --- |
| <!-- uncertain passage -->  <u who="#SPK0" start="#T4" end="#T6">  <w>you</w>  <unclear reason="background noise">  <w>should</w>  </unclear>  <w>let</w>  <!-- [...] -->  </u>  <!-- uncertain passage with alternatives for a single word-->  <u who="#SPK0" start="#T4" end="#T6">  <w>you</w>  <unclear>  <choice>  <w>should</w>  <w>could</w>  </choice>  </unclear>  <w>let</w>  <!-- [...] -->  </u>  <!-- uncertain passage with alternatives for a sequence of words-->  <u who="#SPK0" start="#T4" end="#T6">  <w>I</w>  <w>kiss</w>  <unclear>  <choice>  <seg>  <w>the</w>  <w>sky</w>  </seg>  <seg>  <w>this</w>  <w>guy</w>  </seg>  </choice>  </unclear>  <w>let</w>  <!-- [...] -->  </u>  <!-- incomprehensible passage within an utterance -->  <u who="#SPK0" start="#T4" end="#T6">  <w>good</w>  <w>morning</w>  <gap reason="incomprehensible" unit="syllables" quantity="2"/>  </u>  <!-- incomprehensible passage between utterances -->  <!-- with start and end attributes -->  <u who="#SPK0" start="#T4" end="#T6">  <w>good</w>  <w>morning</w>  </u>  <gap reason="incomprehensible" dur="PT8.9S"  start="#T6" end="#T7"/>  <!-- omitted passage -->  <gap reason="omission, irrelevant sideline of the conversation" dur="PT8.9S"  start="#T6" end="#T7"/> |

# 6.6 Units above the token and below the <u> level (<seg>)

### 6.6.1 Characterisation

In many transcription systems, speakers’ utterances can be subdivided into chunks comprising more than one token and/or pauses and/or non-audible speech events. Often, these are the “sentence equivalents” of spoken language. If and how these chunks are defined, distinguished and delimited varies greatly between different conventions and is hotly debated. Two popular approaches are the use of pragmatic and syntactic criteria which, for instance, lead to the notion of an utterance (not to be confused with TEI’s definition of an utterance) in the CHAT and HIAT systems, and the use of prosodic criteria which lead to the notion of an intonation phrase in the GAT and DT systems. If such divisions are provided, they are usually intended to be exhaustive and unique, i.e. every element of the utterance is part of one and only one such chunk.

### 6.6.2 Representation as <seg>

Divisions of a **<u>** into smaller segments should be represented by **<seg>** elements. The **@type** attribute should be used to denote the general name of the entity (such as “utterance”, “intonation phrase”). A **@subtype** attribute can be added to provide an additional subclassification. An **@xml:id** attribute can be provided to make the entity addressable for standoff annotation.

### 6.6.3 Further constraints

Nesting of **<seg>** elements is possible in principle, but does not occur in most transcription systems. In legacy systems, punctuation (see section 6.4) is often used to delimit and characterise these units.

### 6.6.4 Examples

|  |
| --- |
| <!-- u divided into two seg elements (utterances according to HIAT/CHAT) -->  <u who="#SPK0" start="#T40" end="#T43">  <seg type="utterance" subtype="declarative" xml:id="seg23">  <w xml:id="w319">And</w>  <gap reason="incomprehensible"/>  <w xml:id="w320">disappointed</w>  <w xml:id="w321">when</w>  <w xml:id="w322">you</w>  <w xml:id="w323">got</w>  <w xml:id="w324">to<anchor synch="#T41"/>gether</w>  </seg>  <anchor synch="#T42"/>  <seg type="utterance" subtype="interrogative" xml:id="seg24">  <gap type="incomprehensible"/>  <w xml:id="w325">you</w>  <pc>,</pc>  <w xml:id="w326">Victoria</w>  </seg>  </u>  <!-- u divided into two seg elements (intonation phrases according to GAT/DT) -->  <u who="#SPK0" start="#T40" end="#T43">  <seg type="intonation-phrase" subtype="rising">  <w xml:id="w319">And</w>  <gap reason="incomprehensible"/>  <w xml:id="w320">disappointed</w>  <w xml:id="w321">when</w>  <w xml:id="w322">you</w>  <w xml:id="w323">got</w>  <w xml:id="w324">to<anchor synch="#T41"/>gether</w>  </seg>  <anchor synch="#T42"/>  <seg type="intonation-phrase" subtype="high-rising">  <gap reason="incomprehensible"/>  <w xml:id="w325">you</w>  <pc>,</pc>  <w xml:id="w326">Victoria</w>  </seg>  </u> |

# 7 Bibliographical reference

**Barras, Claude; Geoffrois, Edouard; Wu, Zhibiaou & Liberman, Mark (2000).** Transcriber: Development and Use of a Tool for Assisting Speech Corpora Production. Speech Communication (33), 5-22.

**Bird, Steven & Liberman, Mark (2001).** A formal framework for linguistic annotation. In: Speech Communication (33), 23-60.

**Boersma, Paul & Weenink, David (1996).** PRAAT, a system for doing phonetics by computer, version 3.4. Institute of Phonetic Sciences of the University of Amsterdam, Report 132. 182 pages.

**Broeder, Dan & Wittenburg, Peter (2002).** Management of Language Resources using Metadata. Proceedings of LREC 2002.

**Broeder, Dan & van Uytvanck, Dieter (2014).** Metadata Formats. In: Durand, Jacques; Gut, Ulrike & Kristoffersen, Gjert (eds.): The Oxford Handbook of Corpus Phonology, pp. 150-165.

**DuBois, John; Schuetze-Coburn, Stephan; Cumming, Susanna & Paolino, Danae (1993).** Outline of Discourse Transcription. In: Edwards, Jane & Lampert, Martin (eds.): Talking Data: Transcription and Coding in Discourse Research, pp. 45–89. Hillsdale, NJ: Erlbaum.

**Kipp, Michael (2014).** ANVIL: The Video Annotation Research Tool. In: Durand, Jacques; Gut, Ulrike & Kristoffersen, Gjert (eds.): The Oxford Handbook of Corpus Phonology.

**MacWhinney, Brian (2000).** The CHILDES Project: Tools for Analyzing Talk. Mahwah, NJ: Erlbaum. <http://childes.psy.cmu.edu/manuals/chat.pdf>.

**Parisse, Christophe & Morgenstern, Aliyah (2010).** A multi-software integration platform and support for multimedia transcripts of language. LREC 2010 : Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality, May 2010, La Valette, Malta. LREC 2010 Proceedings : Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality, pp. 106-110

**Rehbein, Jochen; Schmidt, Thomas; Meyer, Bernd; Watzke, Franziska & Herkenrath, Annette (2004).** Handbuch für das computergestützte Transkribieren nach HIAT. In: Arbeiten zur Mehrsprachigkeit, Series B (56).

**Romary, Laurent & Witt, Andreas (2014)**. Data formats for phonological corpora. In: Jacques Durand, Ulrike Gut, and Gjert Kristoffersen (eds.): The Oxford Handbook of Corpus Phonology, pp. 166-192.

**Schmidt, Thomas (2005).** Computergestützte Transkription - Modellierung und Visualisierung gesprochener Sprache mit texttechnologischen Mitteln. Frankfurt a. M.: Peter Lang.

**Schmidt, Thomas; Duncan, Susan; Ehmer, Oliver; Hoyt, Jeffrey; Kipp, Michael; Magnusson, Magnus; Rose, Travis & Sloetjes, Han (2009).** An Exchange Format for Multimodal Annotations. In: Michael Kipp, Jean-Claude Martin, P. P. & Heylen, D. (eds.): Multimodal Corpora, Lecture Notes in Computer Science 207-221. Springer.

**Schmidt, Thomas; Elenius, Kjell & Trilsbeek, Paul (2010).** Multimedia Corpora (Media encoding and annotation). Draft submitted to CLARIN WG 5.7. as input to CLARIN deliverable D5.C-3 "Interoperability and Standards" [http://www.clarin.eu/system/files/clarin-deliverable-D5C3\_v1\_5-finaldraft.pdf]

see also: <http://www1.uni-hamburg.de/exmaralda/files/CLARIN_Standards.pdf>

**Schmidt, Thomas & Schütte, Wilfried (2010)** FOLKER: An Annotation Tool for Efficient Transcription of Natural, Multi-party Interaction. In: Calzolari, N, Choukri, K. et al. (eds.): Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta: European Language Resources Association (ELRA).

**Schmidt, Thomas (2011).** A TEI-based Approach to Standardising Spoken Language Transcription. Journal of the Text Encoding Initiative [Online], Issue 1 | June 2011, URL : [[http://jtei.revues.org/142]](http://jtei.revues.org/142) ; DOI : 10.4000/jtei.142

**Schmidt, Thomas & Wörner, Kai (2014).** EXMARaLDA. In: Durand, Jacques; Gut, Ulrike & Kristoffersen, Gjert (eds.): The Oxford Handbook of Corpus Phonology, pp. 402-419.

**Selting, Margret et al. (2009).** Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion 10 (2009), pp. 353-402.

**Sloetjes, Han (2014).** ELAN: Multimedia Annotation Application. In: Durand, Jacques; Gut, Ulrike & Kristoffersen, Gjert (eds.): The Oxford Handbook of Corpus Phonology, pp. 305-320.

# Annex A

## TEI-ODD specification

**N.B.: This section is not up-to-date. An updated ODD specification can be provided by Lou Burnard. Source for this is currently maintained at https://code.google.com/p/tei-fr/source/browse/trunk/Projects/ISOmulti/ISOmulti.odd**

|  |
| --- |
| <schemaSpec ident="ISOspoken" start="TEI teiCorpus">  <moduleRef key="tei"/>  <moduleRef key="header" except="biblFull"/>  <moduleRef key="core" include="media abbr desc unclear choice ab teiCorpus" />  <moduleRef key="textstructure" include="TEI text div body"/>  <moduleRef key="namesdates" include="persName"/>  <moduleRef key="spoken" include="pause vocal kinesic incident  u shift writing"/>  <moduleRef key="linking" include="ab seg when timeline anchor"/>  <moduleRef key="analysis" include="pc span spanGrp w"/>  <moduleRef key="corpus"/>    <!-- new element for documenting transcription conventions -->  <elementSpec ident="transcriptionDesc"  ns="http://iso-tei-spoken.org/ns/1.0" >  <desc>describes the set of transcription conventions used</desc>  <classes>  <memberOf key="att.global"/>  <memberOf key="model.encodingDescPart"/>  </classes>  <content>  <group xmlns="http://relaxng.org/ns/structure/1.0">  <oneOrMore>  <ref name="model.labelLike"/>  </oneOrMore>  <choice>  <zeroOrMore>  <ref name="model.ptrLike"/>  </zeroOrMore>  <zeroOrMore>  <ref name="model.pLike"/>  </zeroOrMore>  </choice>  </group>  </content>  <attList>  <attDef ident="ident" usage="req">  <desc>supplies an identifier for the encoding convention, independent of any version number.</desc>  <datatype>  <ref xmlns="http://relaxng.org/ns/structure/1.0" name="data.name"/>  </datatype>  </attDef>  <attDef ident="version" usage="opt">  <desc>supplies a version number for the encoding conventions  used, if any.</desc>  <datatype>  <ref xmlns="http://relaxng.org/ns/structure/1.0" name="data.versionNumber"/>  </datatype>  </attDef>  </attList>  <exemplum xml:lang="en">  <egXML xmlns="http://www.tei-c.org/ns/Examples">  <transcriptionDesc ident="HIAT" version="2004"/>  </egXML>  </exemplum>  </elementSpec>    <!-- new element for grouping annotation and utterance -->    <elementSpec ident="annotatedU" ns="http://iso-tei-spoken.org/ns/1.0">  <desc>groups an utterance with the annotation layers associated with  it</desc>  <classes>  <memberOf key="model.divPart.spoken"/>  </classes>  <content>  <group xmlns="http://relaxng.org/ns/structure/1.0">  <ref name="u"/>  <oneOrMore>  <ref name="spanGrp"/>  </oneOrMore>  </group>  </content>  </elementSpec>    <!-- attributes for synchronization -->    <elementSpec ident="when" module="linking" mode="change">  <attList>  <attDef ident="xml:id" mode="change" usage="req"/>  </attList>  <!-- a schematron constraint shd be added to ensure values for  @absolute are monotonically increasing -->  </elementSpec>      <!-- remove @trans from <u> -->  <elementSpec ident="u" module="spoken" mode="change">  <attList>  <attDef ident="trans" mode="delete"/>  </attList>  <!-- need constraint to say that if @start and @end are missing then  nested <anchor>s must be present -->  </elementSpec>      <!-- make @start and @end obligatory on <kinesic> and <incident> -->  <elementSpec ident="kinesic" module="spoken" mode="change">  <attList>  <attDef ident="start" mode="change" usage="req"/>  <attDef ident="end" mode="change" usage="req"/>  </attList>  </elementSpec>    <elementSpec ident="incident" module="spoken" mode="change">  <attList>  <attDef ident="start" mode="change" usage="req"/>  <attDef ident="end" mode="change" usage="req"/>  </attList>  </elementSpec>      <!-- simplify global attributes -->  <classSpec ident="att.global.linking" type="atts" mode="change">  <attList>  <attDef ident="corresp" mode="delete"/>  <attDef ident="sameAs" mode="delete"/>  <attDef ident="copyOf" mode="delete"/>  <attDef ident="next" mode="delete"/>  <attDef ident="prev" mode="delete"/>  <attDef ident="exclude" mode="delete"/>  <attDef ident="select" mode="delete"/>  </attList></classSpec>    <classSpec ident="att.global.analytic" type="atts" mode="delete"/>  <classSpec ident="att.global.facs" type="atts" mode="delete"/>  <classSpec ident="att.global.change" type="atts" mode="delete"/>    <!-- and remove some other attribute classes too -->    <classSpec ident="att.declaring" type="atts" mode="delete"/>  <classSpec ident="att.datable" type="atts" mode="delete"/>  </schemaSpec> |

# Annex B

## Fully encoded example

|  |
| --- |
| <?xml version="1.0" encoding="UTF-8"?>  <TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:tei="http://www.tei-c.org/ns/1.0">  <teiHeader>  <fileDesc>  <titleStmt>  <title/>  </titleStmt>  <publicationStmt>  <authority><!--Fill me in--></authority>  <availability>  <licence target="someurl"/>  <p><!--Fill me in--></p>  </availability>  <distributor><!--Fill me in--></distributor>  <address><!--Fill me in--></address>  </publicationStmt>  <sourceDesc>  <recordingStmt>  <recording type="video">  <media mimeType="video/xxx"  url="file:/S:/Korpora/EXMARaLDA-Demokorpus/Beckhams/Beckhams.mpg"/>  <media mimeType="audio/xxx"  url="file:/S:/Korpora/EXMARaLDA-Demokorpus/Beckhams/Beckhams.wav"/>  <broadcast>  <ab><!--Fill me in--></ab>  </broadcast>  <equipment>  <ab><!--Fill me in--></ab>  <ab><!--Fill me in--></ab>  </equipment>  </recording>  </recordingStmt>  </sourceDesc>  </fileDesc>  <profileDesc>  <particDesc>  <person xml:id="SPK0" n="PAR" sex="1">  <persName/>  </person>  <person xml:id="SPK1" n="VIC" sex="2">  <persName/>  </person>  <person xml:id="SPK2" n="DAV" sex="1">  <persName/>  </person>  </particDesc>  <settingDesc>  <place><!--Fill me in--></place>  <setting>  <activity><!--Fill me in--></activity>  </setting>  </settingDesc>  </profileDesc>  <encodingDesc>  <appInfo>  <application ident="EXMARaLDA" version="1.5.3">  <label>EXMARaLDA Partitur-Editor</label>  <desc>Transcription Tool providing a TEI Export</desc>  </application>  </appInfo>  <transcriptionDesc ident="HIAT" version="2004">  <p>Halbinterpretative Arbeitstranskription</p>  </transcriptionDesc>  </encodingDesc>  <revisionDesc>  <change when="2014-06-23T11:05:11.237+02:00">  Created by XSL transformation from  an EXMARaLDA basic transcription  </change>  </revisionDesc>  </teiHeader>  <text>  <timeline unit="s" origin="#T0">  <when absolute="00:00:00.0" xml:id="T0"/>  <when xml:id="T1" interval="2.1866329681774834" since="#T0"/>  <when xml:id="T2" interval="2.4399623974175575" since="#T0"/>  <when xml:id="T3" interval="2.706624954512373" since="#T0"/>  <when xml:id="T4" interval="3.746608927182151" since="#T0"/>  <when xml:id="T5" interval="4.713260696650855" since="#T0"/>  <when xml:id="T6" interval="5.0732551487288555" since="#T0"/>  <when xml:id="T7" interval="7.586549749347489" since="#T0"/>  <when xml:id="T8" interval="8.533201827034082" since="#T0"/>  <when xml:id="T9" interval="11.366491496166491" since="#T0"/>  </timeline>  <body>  <annotatedU who="#SPK0" start="#T0" end="#T9" xml:id="au1">  <u xml:id="u1">  <seg xml:id="seg0" type="utterance" subtype="declarative">  <w xml:id="w1">And</w>  <w xml:id="w2">what</w>  <w xml:id="w3">comes</w>  <unclear>  <w xml:id="w4">to</w>  <w xml:id="w5">as</w>  <w xml:id="w6">your</w>  </unclear>  <w xml:id="w7">determination</w>  <anchor synch="#T1"/>  <w xml:id="w8">at</w>  <anchor synch="#T2"/>  <w xml:id="w9">all</w>  <anchor synch="#T3"/>  <w xml:id="w10">cost</w>  <w xml:id="w11">to</w>  <w xml:id="w12">actually</w>  <anchor synch="#T4"/>  <pause type="medium"/>  <w xml:id="w13">succeed</w>  </seg>  <anchor synch="#T5"/>  <seg xml:id="seg1" type="utterance" subtype="interrogative">  <w xml:id="w14">I</w>  <w xml:id="w15">mean</w>  <anchor synch="#T6"/>  <w xml:id="w16">is</w>  <w xml:id="w17">that</w>  <w type="repair" xml:id="w18">a</w>  <w xml:id="w19">sort</w>  <w xml:id="w20">of</w>  <w xml:id="w21">a</w>  <w xml:id="w22">message</w>  <w xml:id="w23">that</w>  <w xml:id="w24">you</w>  <w xml:id="w25">hope</w>  <w xml:id="w26">comes</w>  <w xml:id="w27">across</w>  <w xml:id="w28">to</w>  <anchor synch="#T7"/>  <pause dur="PT0.4S"/>  <w xml:id="w29">to</w>  <w xml:id="w30">kids</w>  </seg>  <anchor synch="#T8"/>  <seg xml:id="seg2" type="utterance" subtype="interrogative">  <w xml:id="w31">Because</w>  <w xml:id="w32">a</w>  <w xml:id="w33">lot</w>  <w xml:id="w34">of</w>  <w xml:id="w35">kids</w>  <w xml:id="w36">think</w>  <w xml:id="w37">that</w>  <w xml:id="w38">people</w>  <w xml:id="w39">just</w>  <w xml:id="w40">become</w>  <w xml:id="w41">famous</w>  <w xml:id="w42">over</w>  <w xml:id="w43">night</w>  <pc>,</pc>  <w xml:id="w44">don't</w>  <w xml:id="w45">they</w>  </seg>  </u>  </annotatedU>  <annotatedU who="#SPK1" start="#T2" end="#T3" xml:id="au1">  <u xml:id="u1">  <seg xml:id="seg37" type="utterance" subtype="modeless">  <w xml:id="w46">Yeah</w>  </seg>  </u>  <spanGrp type="nv">  <span from="#T2" to="#T3">nods</span>  </spanGrp>  </annotatedU>  <annotatedU who="#SPK1" start="#T5" end="#T6" xml:id="au1">  <u xml:id="u1">  <seg xml:id="seg38" type="utterance" subtype="modeless">  <w xml:id="w47">Mhm</w>  </seg>  </u>  <spanGrp type="nv">  <span from="#T5" to="#T6">nods</span>  </spanGrp>  </annotatedU>  </body>  </text>  </TEI> |

# 

# Annex C

## Element index

|  |  |
| --- | --- |
| **Name** | **Section(s)** |
| **<anchor>** | 5.2 |
| **<annotationGrp>** | 5.4 |
| **<appInfo>** | 4.3 |
| **<choice>** | 6.5 |
| **<desc>** | 5.5  6.3 |
| **<div>** | 5.7 |
| **<encodingDesc>** | 4.3 |
| **<fileDesc>** | 4.1 |
| **<gap>** | 6.5 |
| **<incident>** | 5.5  6.3 |
| **<kinesic>** | 6.3 |
| **<media>** | 4.1.2 |
| **<particDesc>** | 4.2.1 |
| **<pause>** | 5.5  6.2 |
| **<pc>** | 6.4 |
| **<person>** | 4.2.1 |
| **<profileDesc>** | 4.2 |
| **<publicationStmt>** | 4.1.1 |
| **<recording>** | 4.1.2 |
| **<recordingStmt>** | 4.1.2 |
| **<seg>** | 6.6 |
| **<settingDesc>** | 4.2.2 |
| **<shift>** | 5.6 |
| **<span>** | 5.3 |
| **<spanGrp>** | 5.3 |
| **<teiHeader>** | 4 |
| **<timeline>** | 5.1 |
| **<transcriptionDesc>** | 4.3 |
| **<u>** | 5.2 |
| **<unclear>** | 6.5 |
| **<vocal>** | 6.3 |
| **<w>** | 6.1 |
| **<when>** | 5.1 |

## Attribute index

|  |  |  |
| --- | --- | --- |
| **Name** | **Parent** | **Section** |
| **@absolute** | **<when>** | 5.1 |
| **@ana** | **<w>** | 6.1 |
| **@dur** | **<pause>** | 5.5  6.2 |
| **@dur** | **<gap>** | 6.5 |
| **@end** | **<u>** | 5.2 |
| **@end** | **<annotationGrp>** | 5.4 |
| **@end** | **<incident>** | 5.5  6.3 |
| **@end** | **<pause>** | 5.5 |
| **@feature** | **<shift>** | 5.6 |
| **@from** | **<span>** | 5.3 |
| **@interval** | **<when>** | 5.1 |
| **@lemma** | **<w>** | 6.1 |
| **@lemmaRef** | **<w>** | 6.1 |
| **@mimeType** | **<media>** | 4.1.2 |
| **@n** | **<person>** | 4.2.1 |
| **@new** | **<shift>** | 5.6 |
| **@origin** | **<timeline>** | 5.1 |
| **@reason** | **<unclear>** | 6.5 |
| **@reason** | **<gap>** | 6.5 |
| **@rend** | **<pause>** | 6.2 |
| **@since** | **<when>** | 5.1 |
| **@start** | **<u>** | 5.2 |
| **@start** | **<annotationGrp>** | 5.4 |
| **@start** | **<incident>** | 5.5  6.3 |
| **@start** | **<pause>** | 5.5 |
| **@subtype** | **<div>** | 5.7 |
| **@subtype** | **<seg>** | 6.6 |
| **@synch** | **<shift>** | 5.6 |
| **@synch** | **<anchor>** | 5.2 |
| **@to** | **<span>** | 5.3 |
| **@trans** | **<u>** | 5.2 |
| **@type** | **<recording>** | 4.1.2 |
| **@type** | **<u>** | 5.2 |
| **@type** | **<spanGrp>** | 5.3 |
| **@type** | **<pause>** | 5.5  6.2 |
| **@type** | **<div>** | 5.7 |
| **@type** | **<w>** | 6.1 |
| **@type** | **<pc>** | 6.4 |
| **@type** | **<seg>** | 6.6 |
| **@unit** | **<timeline>** | 5.1 |
| **@unit** | **<pc>** | 6.4 |
| **@url** | **<media>** | 4.1.2 |
| **@who** | **<u>** | 5.2 |
| **@who** | **<annotationGrp>** | 5.4 |
| **@xml:id** | **<person>** | 4.2.1 |
| **@xml:id** | **<when>** | 5.1 |
| **@xml:id** | **<u>** | 5.2 |
| **@xml:id** | **<annotationGrp>** | 5.4 |
| **@xml:id** | **<w>** | 6.1 |
| **@xml:id** | **<seg>** | 6.6 |
| **@xml:lang** | **<spanGrp>** | 5.4 |