New work item proposal

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

# 

# 

# Table of Contents

[Table of Contents](#h.hagawfj907ym)

[Experts](#h.7degqrm52wou)

[Introduction](#h.yysh5i7gq8pk)

[1 Scope](#h.qyn0r4puh88a)

[2 Normative references](#h.yfg9qbw4veac)

[3 Terms and definitions](#h.k8lhntxmgnlv)

[4 General document structure](#h.mxd7fdb9hf8f)

[5 Metadata](#h.tp11p16vrk1o)

[5.1. Description of the electronic file (<fileDesc>)](#h.c4bc0jecfh5n)

[5.1.1. Distribution information (<publicationStmt>)](#h.k3tmktpev7ms)

[5.1.2 Recording information (<recordingStmt>)](#h.u8r4mr3tw2on)

[5.2. Description of circumstances (<profileDesc>)](#h.7e144u94vjs1)

[5.2.1. Participant information (<particDesc>)](#h.m6wjj4g2ebgj)

[5.2.2. Setting information (<settingDesc>)](#h.jkdr948sajnl)

[5.3. Description of source (<encodingDesc>)](#h.g58o0g7l3can)

[6 Macrostructure](#h.jokwrjg8j1vs)

[6.1 Timeline (<timeline>)](#h.dfbr5b22grbz)

[6.2 Utterances (<u>)](#h.31tvgqh77j5)

[6.3 Free dependent annotations (<spanGrp>, <span>)](#h.4v12ridd6jc5)

[6.4 Grouping of utterances and dependent annotations (<annotatedU>)](#h.80clzsvnajnb)

[6.5 Independent elements outside utterances (<pause> and <incident>)](#h.dfbr5b22grbz)

[6.6 Inline paralinguistic annotation (<shift>)](#h.m6zjz6nr4b4n)

[6.7 Global divisions of a transcription (<div>)](#h.vvqfk2pnairk)

[7 Microstructure](#h.cch0wsj32244)

[7.1 Words (<w>)](#h.lxgbc8fwubin)

[7.2 Pauses (<pause>)](#h.z3o3yr42f496)

[7.3 Audible and visible non-speech events (<vocal>, <kinesic> and <incident>)](#h.bbkr4lki56qj)

[7.4 Punctuation (<pc>)](#h.vdzrkhs7l0s2)

[7.5 Uncertainty, alternatives, incomprehensible and omitted passages (<unclear>, <choice>, <gap>)](#h.19xwnk4qpcrs)

[7.6 Units above the word and below the <u> level (<seg>)](#h.gvsavu68whzo)

[8 Bibliographical reference](#h.mr3zkmijs0p1)

[9 Annex](#h.vm9jzd9qo1ot)

[10 Annex - fully encoded example](#h.yaezeltxcvgx)

# 

# Experts

Core Group:

* Lou Burnard
* Bertrand Gaiffe
* Laurent Romary
* Thomas Schmidt
* Andreas Witt

EIT / MMI Group:

* Nadia Mana from FBK (Trento, Italy)
* Tatjana Scheffler (DFKI, Germany)
* Khiet Truong (Univ of Twente)
* Benjamin Weiss (TU Berlin)
* Mathias Wilhelm (DAI Labor)
* Felix Burkhardt (Deutsche Telekom's Innovation Lab)

Working group 'Interopérabilité' of the Consortium ‘Corpus Oraux et multimodaux (IRCOM)’

* Coordinators: Christophe Parisse, Carole Etienne, Bernard Bel.
* Participants : Sarra El Ayari, Nicolas Ballier, Christophe Benzitoun, Philippe Blache, Annelies Braffort, Christian Chanard, Dominique Fohr, Séverine Guillaume, Clément Planck.

International experts (delegates)

* to be added by Andreas Witt

Tool developers (not yet contacted):

* ~~Michael Kipp (FH Augsburg, developer of ANVIL)~~
* ~~Han Sloetjes (MPI Nijmegen, developer of ELAN)~~
* Kai Wörner (Hamburg, co-developer of EXMARaLDA)
* ~~Transcriber: Edouard Geoffrois, Claude Barras~~

~~Other experts:~~

* Daniel Jettka, Hanna Hedeland (Hamburg Centre for Language Corpora)
* ~~Ulrike Gut (Münster, Corpus Phonology Network)~~
* ~~Florian Schiel, Christoph Draxler (München, Bavarian Archive for Sound Signals)~~

# 

# Foreword

**“The foreword of ISO documents are standard texts drafted by ISO’s Technical Management Board. ISO Central Secretariat (ISO/CS) inserts them during editing 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. Typically, transcription of spoken language here 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 be relevant also 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)

The standard aims at being 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.

It 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 regards 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 conformant to the TEI based representation of tokens ISO 24611 (MAF).

# 

# 

# 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 standard makes use of 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)”.

# 

# 

# 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

**POS** Part of speech

**TEI** Text Endcoding Initiative

# 

# 

# 

# 4 Metadata

The TEI Guidelines formulate extensive suggestions for encoding metadata inside different subsections of the **<teiHeader>** element. In the following, only those pieces of metadata are addressed 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">  <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>BBC studio London</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">  <p>Orthographic transcription according to HIAT</p>  </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, i.e. 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 **<annotatedU>** element (see below), **<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 **<annotatedU>** element (see below), **<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 7.

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 be represented inside **<u>** elements, and the other one be represented in appropriate **<span>** elements (see below).

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

Whereas **<u>** contains the basic (typically, but not necessarily) 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 two 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 **<annotatedU>** element (see below).

On the level of tokens, annotations via **<span>** elements pointing to **<w>** elements is conformant 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.

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> |

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

**<u>** elements and the annotations referring to them can be grouped under an **<annotatedU>** 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 **<annotatedU>** is optional, it is not allowed to mix **<annotatedU>** and **<u>** elements on the top level - i.e. as soon as one **<annotatedU>** element is used, all **<u>** elements have to be wrapped inside an **<annotatedU>** element.

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

If **<annotatedU>** 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 **<annotatedU>** addressable for stand-off annotations.

|  |
| --- |
| <!-- an utterance grouped with corresponding annotations -->  <annotatedU 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>  </annotatedU>  <!-- an annotatedU without subordinate <u> element -->  <annotatedU who="#SPK0" start="#T0" end="#T1">  <vocal>  <desc>laughter</desc>  </vocal>  </annotatedU> |

## 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 **<annotatedU>** (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.

|  |
| --- |
| <annotatedU who="#SPK0" start="#T0" end="#T1">  <!-- [...] u and spanGrp elements, see above -->  </annotatedU>  <!-- 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"/>  <annotatedU who="#SPK1" start="#T3" end="#T4">  <!-- [...] u and spanGrp elements, see above -->  </annotatedU> |

## 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 above) 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> -->  <annotatedU 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>  </annotatedU> |

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

For a division of a transcription into larger sections (above the level of **<u>** or **<annotatedU>** 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">  <annotatedU who="#SPK0" start="#T0" end="#T1">  <!-- [...] u and spanGrp elements, see above -->  </annotatedU>  <annotatedU who="#SPK1" start="#T1" end="#T2">  <!-- [...] u and spanGrp elements, see above -->  </annotatedU>  </div>  <!-- main part, not embedded in a div element -->  <annotatedU who="#SPK0" start="#T2" end="#T3">  <!-- [...] u and spanGrp elements, see above -->  </annotatedU>  <!-- [...] -->  <!-- final section of the interaction -->  <div type="farewell">  <annotatedU who="#SPK1" start="#T112" end="#T113">  <!-- [...] u and spanGrp elements, see above -->  </annotatedU>  <annotatedU who="#SPK0" start="#T113" end="#T114">  <!-- [...] u and spanGrp elements, see above -->  </annotatedU>  </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 above). 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 below.

### 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 -->  <annotatedU 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>  <annotatedU>  <!-- 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 above) 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 the most common such phenomena described in transcriptions 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 **<annotatedU>**, 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 **<annotatedU>** 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 annotatedU -->  <!-- with start and end points -->  <annotatedU 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>**  </annotatedU>  <!-- (backchannel) nodding encoded as kinesic element on the level of annotatedUs -->  <!-- with speaker assignment and start and end points -->  <annotatedU who="#SPK0" start="#T6" end="#T9">  <!-- [...] -->  </annotatedU>  <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 punctutation 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 above) or a **@type** attribute on a **<seg>** element (for tone movements, see 7.6.). 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 for describing uncertainty in transcription, i.e. parts where the transcriber is not sure what he heard, and for 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 also 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 **<annotatedU>** 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. How (and if) 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 7.4.) 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>  </div>  <!-- 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>  </div> |

# 

# 8 Bibliographical reference

**Barras, C.; Geoffrois, E.; Wu, Z. & Liberman, M. (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, P. & Weenik, D. (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, D. & Wittenburg, P. (2002).** Management of Language Resources using Metadata. Proceedings of LREC 2002.

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

**DuBois, J., S. Schuetze-Coburn, , S. Cumming, and D. Paolino (1993).** Outline of Discourse Transcription. In Talking Data: Transcription and Coding in Discourse Research, ed. J. Edwards and M. Lampert, 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, B. (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, J.; Schmidt, T.; Meyer, B.; Watzke, F. & Herkenrath, A. (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, T.; Duncan, S.; Ehmer, O.; Hoyt, J.; Kipp, M.; Magnusson, M.; Rose, T. & Sloetjes, H. (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, T. & Schütte, W. (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, M. 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.**

An ODD for the proposals might look like this.

|  |
| --- |
| <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> |

Source for this is currently maintained at <https://code.google.com/p/tei-fr/source/browse/trunk/Projects/ISOmulti/ISOmulti.odd>

# 

# 

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

## Attribute index

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

## 

## 

## Minutes ISO/DIN Meeting Berlin, 22-October-2012

(Lou Burnard, see also <http://www.tei-c.org/Activities/Council/Working/tcw25.xml>)

EIT MMI Meeting, Berlin 22 oct 2012  
  
As noted at the last FTF, Laurent Romary in his capacity as ISO TC7 WG3 chair has proposed a new ISO/TEI joint activity in the area of speech transcription, which comes with the slightly obscure label of EIT MMI: the last part of which is short for “multimodal interaction”, although it seems the activity is really only concerned with speech transcription. I was invited to attend the third EIT MMI workshop, held at the DIN's offices in Berlin.

Prime movers in the activity, apart from Laurent, appear to be Thomas Schmidt and Andreas Witt from the Institut fur Deutsche Sprache in Mannheim, but a number of other European research labs, mostly concerned with analysis of corpora of human computer interaction, were also represented; specifically: Nadia Mana from FBK (Trento, Italy); Tatjana Scheffler (DFKI, Germany); Khiet Truong (Univ of Twente) ; Benjamin Weiss (TU Berlin); Mathias Wilhelm (DAI Labor); Bertrand Gaiffe (ATILF, Nancy). This being an ISO activity, the real world of commerce and industry was also represented by Felix Burkhardt from Deutsche Telekom's Innovation Lab.  
Related ISO activity mentioned by Laurent included the work on Discourse Relations led by Harry Bunt, and the long-awaited MAF (morpho-syntactic annotation framework) which are both due to appear Real Soon Now. A quick tour de table confirmed my impression that most of the attendees were primarily researchers in Human Computer Interaction with little direct experience of the construction or encoding of spoken corpora, but Thomas Schmidt more than made up for that.

The main business of the day was to go through his preliminary draft working document, the objective of which is to confer ISO authority on a subset of the existing TEI proposals for spoken text transcription, with some possible modification. The underlying work is well described in Schmidt's recent excellent article in TEIJ, so I won't repeat it: essentially, it consists of a close look at the majority of transcription formats used by the relevant research community/ies and tools, a synthesis of what they have in common, and suggestions of how that synthesis maps to TEI. This is to a large extent motivated by concerns about preservation and migration of data in “legacy” formats.  
  
The discussion began by establishing boundaries: despite my proposal to the contrary, it seems there was little appetite to extend the work into the area of truly multimodal transcriptions, which was still generally felt to be insufficiently understood for a practice-based standard to be appropriate. Concern was expressed that we should not make ad hoc premature suggestions. So the document really only concerns transcribed   
speech. There was no disagreement with the general approach which is to distinguish a small number of macro-structural features provide guidelines about how to mark up specific units of analysis at the micro-structural level, using a subset of the TEI. I was also much cheered by two further remarks he made the graph-based “annotation framework” formalisation proposed by Bird and Liberman was theoretically complete but so generic as to be practically useless (I paraphrase) at the micro level, everything you need is there in the TEI (I quote)  
  
Discussion focussed on the following points raised by the working document:

* **Tiers:** Many existing tools organise transcriptions into “tiers” of annotation.   
  These seem to be purely technical artefacts, which can be addressed more   
  exactly by used of XML markup. Unlike “levels” of annotation, they have   
  no semantics. It's doubtful that we need a <tier> element.
* **Metadata -1:** How many of the (very rich) TEI proposals should be included, or mentioned? And how should the three things Thomas had found missing be supplied? I suggested that <appinfo> was an appropriate way to record information about the transcription tool used; that the definition of the transcription system used belonged in the <encodingDesc>; and agreed that there was nothing specifically provided for recording pointers or links to the original video or audio transcribed. In the meeting, I   
  speculated that maybe there was scope for extending (or misusing) <facsimile> for this last purpose; another possibility which occurs to me as I type these notes is that one could also extend <recordingDesc>.
* **Timing:** The timeline is fundamental to the macrostructure of a transcript. Thomas' examples all used absolute times for its <when>s, but I suggested that relative ones might be easier. The document ordering both of <when>s and of transcribed speech should reflect the temporal order as far as possible; this would allegedly facilitate interoperability
* **Metadata-2:** What metadata was needed, required, recommended for the description of participants? (@sex raised its ugly head here). Could we use <person> to refer to artificial respondents in MMI experiments? (yes, if they have person-like characteristics; no otherwise).   
  It was noted that almost any personal trait or state might be crucial to the analysis of some corpora. We noted that CMDI now recommended using the ISOCAT data category registry as an independent way of defining metadata terminology; also that ISOCAT was now available within the TEI scheme (though whether it fits into personal metadata I am less sure). There was (I think) general agreement that we'd reference the various options available in the TEI but not incorporate all of them. We agreed that the principles underlying a given transcription should be clearly documented, either in associated articles, in the formal specification for an encoding, or in the header of individual documents.
* **Utterances:** Several people disliked the expanded element name <u> and its definition, for various theoretical reasons. Its definition should be modified to remove the implication that it necessarily followed a silence, though we seemed to agree that a <u> could only contain a stretch of speech from a single speaker.  
  The temporal alignment of a <u> can be indicated either by @start and @end or by nested <anchor/>s : the standard should probably recommend use of one or the other methods but not both. We discussed whether or not the fact that existing tools did not support the (even simpler) use of @trans to indicate overlap should lead us not to recommend it.
* **U-plus:** Thomas wanted some method of associating with a <u> the whole block of   
  annotations made on it (represented as one or more <interpGrp>s). His document suggested using <div> for this purpose. A lighter-weight solution might be to include <interpGrp> within <u>, or to propose a new wrapper <annotatedU> element.
* **Tokenization:** Laurent noted that MAF recommended use of <w> for individual tokens; we didn't need to take a stand on the definition of “word” but could simply refer to MAF. We needed some way of signalling the things that older transcription formats had found important, e.g. words considered incomplete, false starts, repetitions, abbreviations etc. so we needed to choose an appropriate TEI construct for them, even if we thought the concept was not useful or ill-defined. The general purpose <seg> element might be the simplest solution, but some diplomacy would be needed about how to define its application and its possible @type or @function values.

**Conclusions**  
  
This workgroup will probably produce a useful document describing an important use case for the TEI recommendations on spoken language. It is currently a Google Doc which the group has agreed to share with the Council. I undertook to help turn this into an ODD, which could eventually become one of our Exemplars. Work on standardising other aspects of transcribed multimodal interactions probably needs to be deferred to a later stage.

## Comments Carol Etienne (by e-mail 11-February-2013)

Les membres du groupe de travail 'Interopérabilité' de l'IRCOM, IR Corpus Oraux et Multimodaux, sont intéressés par l'utilisation de la TEI pour les corpus oraux, dans les projets regroupant plusieurs "sources" de corpus ou bien en tant que format d'échange de données entre les logiciels d'annotation comme Transcriber, Clan, Elan ou Anvil majoritairement utilisés dans notre communauté.

Les besoins en terme de métadonnées comme en terme de transcription sont différents et plus ou moins complexes suivant les objets d'étude de chaque équipe mais les besoins suivants émergent :

1) des **métadonnées pour les locuteurs** : données sociolinguistiques, relation entre les utilisateurs, langue maternelle et autres langues, langue des parents, situation professionnelle, locuteurs génériques (client/commerçant, hotline, élève/enseignant, ...) La Tei fournit un ensemble important de balises dans <person> qui devraient couvrir la majorité des besoins

2) des **métadonnées pour l'enregistrement audio ou vidéo** : type, qualité, anonymisation, url, ...

La balise <recording> reste à ma connaissance bien pauvre et <equipment> peu formatée

3) des **métadonnées pour les langues utilisées** : segments clairement identifiées vs segments difficiles à catégoriser, code-switching, langues de contact, ...

Les balises <langUsage> et <language> reposent sur une identification de la langue , une langue principale mais gèrent assez mal les segments non attribués à moins de définir une typologie de langue ...

4) des **métadonnées pour identifier la nature de l'oral**, le nombre de locuteurs, ... : récit, enseignement, apprentissage, conversation, … La balise <setting> peut répondre mais elle est peu formatée

5) dans la transcription, plusieurs niveaux d'annotations sont réalisés et ne correspondent pas forcément au même découpage temporel de cette transcription, le non-verbal n'a par exemple pas les mêmes bornes que le verbal, la traduction s'entend sur un segment de plusieurs mots alors que les lemmes se rapportent à un mot précis, ...

Je pense que nos collègues travaillant sur le syntaxe seront aussi intéressés par un **jeu d'étiquettes liées au mot**.

Pour ce point, les balises <spanGrp> et <anchor> de la TEI peuvent répondre à cette question avec une définition de plusieurs bornes (repère temporel ou lien direct à l'id d'un mot ou un anchor "éléctron libre" sans lien avec le timeline ou le mot) pour identifier la portée de ces annotations.

# Comments by Thomas Schmidt / Points of discussion with the Ortolang group, 03 December 2013

# 1. Scope of the document

## Role of metadata

This ISO/TEI proposal is not meant to do a comprehensive standardisation of metadata structure and vocabulary. Our ambition is limited to prescribing preferred ways of encoding metadata that are crucial to processing the transcription itself. This includes so far:

* the definition of speakers in <person> elements, with IDs, to be referred to by @who attributes in the transcription
* a reference to the recording(s) in a <recording> element
* a description of the tool with which the original transcription was created or from which it was converted to TEI in a <appInfo> element
* likewise: a description of the transcription guidelines used in an element <transcriptionDesc>, newly introduced in the ODD (i.e. not in TEI P5)

If there are further metadata items described in the text or illustrated in the examples, this is because TEI already provides a clear way of using them. They are, however, in no way central to the ISO/TEI proposal.

## Role of annotations

Likewise, we do not want to go into too much depth concerning annotations beyond the primary transcription. <spanGrp> with <span> elements are provided as a generic means of dealing with annotations present in many widely used tools. More sophisticated annotation mechanisms are not in the scope of the ISO document, but are provided, for instance, in the ISO standard for feature structures

## Role of gestures etc.

We state in the introduction that

“Typically, transcription of spoken language here 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.”

“Corpora with gestures only” (see Ortolang comments) are therefore definitely not in the scope of the ISO document. The possibility of catering for purely phonetic transcriptions (i.e. without orthographic transcription),however, should be considered (see below).

# 2. Details

## Header

How to encode a short name for a speaker (e.g. ‘LB’ for ‘Lou Burnard’)? (not yet in the document)

* Lou’s suggestion: @n attribute on a <person> elements - @n is defined as “(number) gives a number (or other label) for an element, which is not necessarily unique within the document”
* Ortolang’s suggestion: <code> element - <code> is defined in the guidelines as “contains literal code from some formal language such as a programming language”
* Comment: the definition of <code> in the guidelines does not seem to fit the way it is used in Ortolang. The @n attribute is closer to the mark, although one would have to use the part of the definition included in parentheses
* My suggestion: use the @n attribute, not the <code> element

## Time representation

Explicit timeline, referred to by elements or put time values directly into elements?

* TEI/ISO’s suggestion: explicit timeline
* Ortolang’s suggestion: directly into elements
* Comment: The two solutions are equivalent as long as there are absolute time values for each timepoint in the transcription. However, it happens that some timepoints only have a relative order, but not an absolute time value. In this case, the second solution will not work, while it is possible to have a timeline in which only some of the included <when> elements have absolute time values
* My suggestion: allow both solutions, make the timeline the preferred one, and say why

## Macro structure of the transcription

### “Global” subdivisions

The TEI/ISO document says as yet nothing about divisions of a transcription above the <u>/<annotatedU> elements (example: greeting, main part, leave taking). Ortolang suggests using (possibly nested) <div> elements for this purpose because this type of information is encoded, for instance, in Transcriber and CHAT documents. I totally agree that this is useful and I do not see any problems it would cause. I suggest we add a suitable paragraph in the ISO/TEI document.

### <u> elements, their internal structure, their place in the document

TEI/ISO’s suggestion:

|  |
| --- |
| <annotatedU>  <!-- the (orthographic) transcribed text from the primary tier -->  <u who="#SPK0" start="#T0" end="#T2">  <!-- character data or further elements like <w> -->  <!-- possibly <anchor> elements for temporal subdivision -->  <!-- possibly <seg> elements for intermediate structuring →  <anchor synch="#T0"/>  <seg function="utterance" type="declarative">  <w>bonjour</w>  <anchor synch="#T1"/>  <w>madame</w>  </seg>  </u>  <!-- additional annotations from a sup (=suprasegmentals) tier -->  <spanGrp type="sup">  <!-- spans with references to timeline or other elements -->  </spanGrp>  <!-- additional annotations from an en (=English translation) tier -->  <spanGrp type="en">  <!-- [...] (see above) -->  </spanGrp>  </annotatedU> |

Ortholang’s suggestion:

|  |
| --- |
| <u who="#SPK0" start="12.4" end="15.3">  <!-- orthographic -->  <ortho>  <!-- <g> elements for intermediate structuring -->  </ortho>  <!-- phonetic -->  <phon><!-- PCDATA / IPA --></phon>  <!-- additional annotations from further tiers -->  <add type="mod"><!-- PCDATA --></add> |

Correspondences

|  |  |  |
| --- | --- | --- |
|  | TEI/ISO | Ortolang |
| top level element | <annotatedU> | <u> |
| orthographic transcription | <u> | <ortho> |
| intermediate structuring of orthographic transcription | <seg> | <g> |
| phonetic transcription | no dedicated element, possible as  <spanGrp type=”pho”> | <phon> |
| further annotation tiers of category X | <spanGrp type=”X”> | <add type=”X”> |

Comments

* I can agree with Ortolang’s criticism that <annotatedU> is impractical as a top level element if it is made optional. I can also agree that properties like speakers (i.e. @who attributes) should be assigned to the top level element (and not to a subordinate <u>) if they pertain to all subordinate elements
* In the ISO/TEI proposal, phonetic transcription is regarded as an optional annotation just like any other annotation whereas orthographic transcription is obligatory. In the Ortolang proposal, by contrast, phonetic transcription is privileged (gets a dedicated element) and can even occur without a corresponding orthographic transcription. It is true that there are purely phonetic transcriptions - I wanted to take them out of the scope of this document, Ortolang wants them included. I would be ready to change my mind in this respect.
* neither <ortho> nor <phon> are defined in the P5 guidelines - if they are needed, we need to add them to the ODD - are there any opinions from the TEI experts on this? Is there maybe a suitable element I have overlooked? Could it be better to make the two variants of the same element, such as <transcription type=”orthographic”> and <transcription type=”phonetic”>?
* <g> is defined in the guidelines as “(character or glyph) represents a glyph, or a non-standard character”. I don’t see how this definition fits the way the element is used in Ortolang
* <add> is defined as “(addition) contains letters, words, or phrases inserted in the source text by an author, scribe, annotator, or corrector”, <spanGrp> is (well…) a group of <span> elements which are defined as “associates an interpretative annotation directly with a span of text”. I think the second definition fits the purpose better, the first being somehow very closely tied to text data.

My Suggestion:

* use <u> and only <u> as the obligatory element for top level structuring (i.e. get rid of <annotatedU> in the TEI proposal)
* use <ortho> and <phon> (or two equivalent new or existing elements) for orthographic and phonetic transcription respectively. Modify the content model of <u> so that it is obligatory to have at least one of the two present (may be awkward to describe in XML schema…)
* use <seg> and only <seg> for intermediate structuring of <ortho> (i.e. get rid of <g> in Ortolang’s proposal?)
* use <spanGrp> with <span> for additional annotations (i.e. get rid of <add> in Ortolang’s proposal?)

Decision in skype conference, 5-12-2013

* keep <annotatedU> as top level element
* allow <u> as top level element as long as there is no need for grouping it with other elements (i.e. for simple, orthography only transcriptions)
* disallow mixing of <annotatedU> and <u> within the same document (or corpus?)
* provide a @type attribute for <u>: if not used, type=”orthographic” is assumed as the default, other possible values are ‘phonemic’ and ‘phonetic’. There are probably more, so the list should be left open
* make <u> optional within <annotatedU> to cater for cases where no genuine transcription can be provided
* allow only one <u> within an <annotatedU> meaning that, in cases of multiple (orthographic, phonetic, …) transcriptions, one of them has to be singled out as primary, the others go into appropriately typed <spanGrp> elements
* may be necessary to specify the exact meaning of <u> in an individual document in the <encodingDesc> of the header
* <seg> instead of <g> is fine for the French group(s)

Further points from the skype conference

* <gap>: difference between conscious decisions not to transcribe (e.g. not interesting, sideline conversation) and impossibility to transcribe (incomprehensible, unclear) must be maintained. Maybe <gap> does not have the right semantics for the latter case
* something about prosodic elements, not discussed due to Skype failure. Carole will add details here or via email
* French group may have needs and solutions that go beyond the scope of this document. A summary of these should be exchanged