INTERNATIONAL STANDARD

ISO 24624

First edition
2016-08-15

Language resource management — Transcription of spoken language

Gestion des ressources linguistiques — Transcription du langage parlé

Reference number
ISO 24624:2016(E)

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword

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

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

The committee responsible for this document is ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 4, *Language resource management*.

Introduction

This document sets out to facilitate 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. body language, gestures and facial expression) may be part of a spoken language transcription, but this document starts from the assumption that the verbal dimension is the primary focus of a spoken language transcription. Likewise, although this document 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.

This document is developed in the context of the joint agreement between ISO and the Text Encoding Initiative (TEI) consortium, and accordingly, its content is also distributed as part of the TEI guidelines.[23]

This document takes into account data models and encoding practices supported by widely used transcription software. More specifically, it builds on several interoperability studies[12],[16],[17],[19] involving the following tools:

— ANVIL[10]

— CLAN[11]

— ELAN[22]

— EXMARaLDA[20]

— FOLKER[18]

— Transcriber[1]

This document was developed to be compatible with the formats produced by these tools. The compatibility may extend to the formats of further labelling tools (e.g. Praat[4] or Wavesurfer, http://www.speech.kth.se/wavesurfer/index2.html), but possibly on a lower level and/or with a requirement to convert these formats to one of the above-mentioned before adding mandatory information (e.g. speaker assignment) using the respective tools.

This document also aims to be usable with widely used transcription systems ("conventions"). However, in a technical sense, compatibility is not easily definable in this area since, unlike the tool formats, most of these systems lack an explicit formalization. The following selection of transcription systems was considered for this document:

— Codes for the Human Analysis of Transcripts (CHAT)[11]

— Discourse Transcription (DT)[7]

— Gesprächsanalytisches Transkriptionssystem (GAT)[21]

— Halbinterpretative Arbeitstranskriptionen (HIAT)[13]

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. However, it should be noted that there are several TEI profiles for the CMDI framework which are related both to each other and to CMDI profiles of other metadata formats (e.g. IMDI) via the ISOCAT registry (see also References [5], [6] and [9]).

This document aims to define both a target format for legacy data conversion and a format suitable for future data processing requirements. The pros and cons of these two demands were carefully weighed up before decisions were taken. 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 committee TC 37/SC 4, this document 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 document also aligns with the mechanism proposed in the TEI guidelines to embed stand-off annotations within a TEI document. In particular, this mechanism contains a generic element (<annotationBlock>) that groups together annotations related to the same linguistic segment; this grouping meets the needs of this document in the case of annotations of <u> elements or its children.

Finally, this document is complementary and does not overlap with the speech and multimodal interaction-related standards developed within the W3C. In particular, it does not deal with speech synthesis as is the case for SSML,[24] nor does it deal with the representation of the semantic interpretation of multimodal utterances as does EMMA.[25]

Language resource management — Transcription of spoken language

1 Scope

This document specifies rules for representing transcriptions of audio- and video-recorded spoken interactions in XML documents based on the guidelines of the TEI. As a secondary objective, the document aims to relate transcribed data with standards for annotated corpora. It is applicable to transcription data for studies in sociolinguistics, conversation analysis, dialectology, corpus linguistics, corpus lexicography, language technology, 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.

Annex A gives a fully encoded example and Annex B provides an element index and an attribute index.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org/obp

3.1
dependent annotation
annotation which does not refer directly to an audio or video recording, but to another annotation, typically an orthographic or phonetic transcription

3.2
milestone element
empty XML element used to indicate a boundary point

3.3
orthographic transcription
representation or modelling of spoken language based on the orthography of the respective language

3.4
paralinguistic feature
feature of spoken language beyond the individual sound(s), such as voice quality, pitch, volume, intonation

3.5
phonetic transcription
representation or modelling of spoken language based on the sound system of the respective language

3.6
spoken language
oral language produced by a person's vocal system

3.7
transcriber
person who carries out the transcription

3.8
transcription
representation or modelling of spoken language by means of written symbols

3.9
transcription system
theoretically founded set of principles and rules detailing what spoken language phenomena are to be transcribed, and how they are to be transcribed

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 (i) crucial for ensuring the interpretability and exchangeability of spoken language transcriptions in general or (ii) likely 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.

EXAMPLE 1 Use of <publicationStmt>

```
<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). Only the actual recording(s), usually digital audio and/or video files, should be described here. General information about the respective interaction which is independent of the recording(s) should be described in the **<settingDesc>** element (see 4.2.2).

A **<media>** element inside a **<recording>** element should be used to refer to the corresponding digital file via a **@url** attribute (see Reference [2]). 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 (see Reference [8]) on the **<media>** element. Where two or more files are derived from the same master recording (e.g. a video file or an extracted audio track), these should be represented as different **<media>** elements inside the same **<recording>** element, rather than as different **<recording>** elements. TEI linking mechanisms, such as **<ref>** or **@corresp**, can be used to describe relationships between different recordings or between recordings and other elements, such as speakers.

EXAMPLE 2 Use of <recordingStmt>

```
<!-- a simple case: one video recording of the entire interaction -->
<!-- and a separate audio file containing the audio track of the video -->
<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 the -->
        <!-- 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>

<!-- a more complex case: two synchronous audio files -->
<!-- each recording one specific speaker -->
<recordingStmt>
    <recording type="audio" xml:id="REC1">
        <media mimeType="audio/wav" url="Victoria.wav"/>
        <equipment>
            <ab>Recorded with a ZOOM H4NSP, external lapel microphone
             clipped to <persName corresp="#SPK1">Victoria Beckham</persName>'s
dress</ab>
            <ab>Synchronized with <ref target="#REC2">David Beckham's record-
ing</ref></ab>
        </equipment>
    </recording>
    <recording type="audio" xml:id="REC2">
        <media mimeType="audio/wav" url="David.wav"/>
        <equipment>
            <ab>Recorded with a ZOOM H4NSP, external lapel microphone
            clipped to <persName corresp="#SPK2">David Beckham</persName>'s
            shirt collar</ab>
            <ab>Synchronized with
```

```
            <ref target="#REC1">Victoria Beckham's recording</ref></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. The use of 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 shall 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 or role in the recorded conversation.

EXAMPLE 3 Use of <particDesc>

```
<particDesc>
   <person xml:id="SPK0" sex="1" n="DS" role="interviewer">
     <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" role="interviewee">
     <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 general information about the setting and circumstances of the interaction. This includes such matters as the place and time, spatial organization

and artefacts of the interaction. Information pertaining to a specific recording of that interaction should not be recorded here, but in the **<recordingStmt>** (see 4.1.2).

EXAMPLE 4 Use of <settingDesc>

```
<settingDesc>
  <place>
    <placeName>BBC studio London</placeName>
  </place>
  <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 is used to record information about the way the TEI encoded text has been derived from a recorded source. This includes information about both the tool which created the transcription inside an **<appInfo>** element and the convention 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.

EXAMPLE 5 Use of <encodingDesc>

```
<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; these points 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 shall be in

the same order as the timepoints they refer to. Specifying an **@interval** attribute is optional, but it is very useful for many processing purposes. Absolute time values in the **@interval** attribute should be given in seconds from the start of the recording with the appropriate number of decimal points. The first **<when>** element in the timeline corresponds to the start time of the transcribed recording. If an absolute value is known for this point in time, it can be encoded in an **@absolute** attribute of the first element and the **<timeline>** element can point to it via an **@origin** attribute. If no absolute value for the start of the recording can be provided, the **@origin** and **@absolute** attributes should be omitted.

EXAMPLE 6 Use of <timeline>

```
<timeline unit="s" origin="#T0">
   <when xml:id="T0" absolute="2009-02-04T20:42:00"/>
   <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>** do not lie within the scope of this document. The TEI definition characterizing a **<u>** as "often preceded by a silence or a change of speaker" should be viewed as a suggestion only. It is therefore 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 it is not wrapped inside an **<annotationBlock>** element (see 5.4), a **<u>** element shall 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 stand-off annotation, for instance, via **** elements (see 5.3).

If it is not wrapped inside an **<annotationBlock>** element (see 5.4), a **<u>** element shall be assigned to the timeline by providing values for the **@start** and **@end** attributes pointing to 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.

In multilingual interactions, it may be necessary to record the language of an utterance. This can be done in an **@xml:lang** attribute of the **<u>** element. Alternatively, the language of an utterance can be treated as an annotation and encoded in a **** element (see 5.3). In cases of interactions where code-switching or similar phenomena occur, it can be preferable to record the language of individual tokens (see 6.1) instead of entire utterances.

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.

EXAMPLE 7 Temporal information for <u> elements

```
<!-- 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>

<!-- u with an attribute for language -->
<u who="#SPK1" start="#T0" end="#T1" xml:id="u2" xml:lang="en">Good morning! </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 Example 7). Further structuring of the content of a **<u>** element (e.g. markup of tokens and pauses) may be carried out via the mechanisms described in Clause 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 realizations, such as "eye dialect", "literary transcription" and "modified orthography". If this is the case, no further specification in the form of a **@notation** 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 **@notation** attribute with an appropriate value.

EXAMPLE 8 Phonetic transcription inside a <u> element

```
<!-- u with phonetic transcription in IPA -->
<u who="#SPK1" start="#T0" end="#T1" notation="phonetic">gʊd 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 **** elements (see 5.3).

5.3 Free dependent annotations (<spanGrp>,)

Whereas **<u>** typically, but not necessarily, contains the basic orthographic transcription, **** elements should be used to represent additional annotations (e.g. part-of-speech tagging, prosodic annotation and 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 shall 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 shall be grouped with the **<u>** element they refer to by using an **<annotationBlock>** element (see 5.4). This is necessary to avoid ambiguities of reference in cases of overlapping speech.

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

Alternatively, annotations of single tokens (e.g. lemmatization and part-of-speech tagging) may be realized as appropriate attributes on **<w>** elements if no structural conflicts between the two levels exist (see 6.1.2).

For annotations with an internal structure, nesting **** 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 does not lie within the scope of this document.

EXAMPLE 9 Use of <spanGrp> and for annotations

```
<!-- 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 (<annotationBlock>)

<u> elements and the annotations referring to them can be grouped under an **<annotationBlock>** 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, that is to say, it provides a "tesselation" of the transcription document. **<spanGrp>** elements in which spans point to the timeline rather than directly to other elements of the transcription shall 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 **<annotationBlock>** is optional, it is not allowed to mix **<annotationBlock>** and **<u>** elements on the top level; in other words, as soon as one **<annotationBlock>** element is used, all **<u>** elements have to be wrapped inside an **<annotationBlock>** element.

<annotationBlock> elements shall not contain more than one **<u>** element. However, there may be cases where it makes sense to use an **<annotationBlock>** as a container only for the description of a non-verbal action of a participant (using one of the elements described in 6.3), without a subordinate **<u>** element.

If **<annotationBlock>** is used, 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 **<annotationBlock>** addressable for stand-off annotations.

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

EXAMPLE 10 Use of <annotationBlock>

```
<!-- an utterance grouped with corresponding annotations -->
<annotationBlock 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 a translation tier -->
   <!-- with an xml:lang attribute capturing the language of the translation -->
   <spanGrp type="translation" xml:lang="en">
        <!-- [...] (see above) -->
   </spanGrp>
</annotationBlock>

<!-- an annotationBlock without subordinate <u> element -->
<annotationBlock who="#SPK0" start="#T0" end="#T1">
   <vocal>
      <desc>laughter</desc>
   </vocal>
</annotationBlock>
```

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 this document, these elements appear on the same hierarchical level as **<annotationBlock>** (or, as the case may be, **<u>**) elements. In order to fit them into the temporal structure, they shall have **@start** and **@end** attributes pointing to the timeline.

EXAMPLE 11 Use of <incident> and <pause> outside utterances

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


5.6 Inline paralinguistic annotation (<shift>)

The TEI guidelines provide 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 shall 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 shall 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 **** element (see 5.3) is the preferable alternative.

EXAMPLE 12 Use of <shift>

```
<!-- 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> -->
<annotationBlock 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>
</annotationBlock>
```

5.7 Global divisions of a transcription ( )

For a division of a transcription into larger sections (above the level of **<u>** or **<annotationBlock>** elements), for example, for different phases of an interaction, the ** ** element can be used. The ** ** element may potentially contain more than a single **<u>** or **<annotationBlock>**. Its **@type** and **@subtype** attributes may be used to categorize the larger units as required. This element is entirely optional, but if it is used, a division shall be indicated for the whole of the transcription, that is to say, every **<annotationBlock>** or **<u>** shall be contained by some ** **.

EXAMPLE 13 Use of

```
<!-- initial section of the interaction -->
<div type="greeting">
    <annotationBlock who="#SPK0" start="#T0" end="#T1">
        <!-- [...] u and spanGrp elements, see above -->
    </annotationBlock>
    <annotationBlock who="#SPK1" start="#T1" end="#T2">
        <!-- [...] u and spanGrp elements, see above -->
    </annotationBlock>
</div>
<!-- main part -->
<div>
    <annotationBlock who="#SPK0" start="#T2" end="#T3">
        <!-- [...] u and spanGrp elements, see above -->
    </annotationBlock>
</div>
<!-- [...] -->
<!-- final section of the interaction -->
<div type="farewell">
    <annotationBlock who="#SPK1" start="#T112" end="#T113">
        <!-- [...] u and spanGrp elements, see above -->
    </annotationBlock>
```

```
    <annotationBlock who="#SPK0" start="#T113" end="#T114">
        <!-- [...] u and spanGrp elements, see above -->
    </annotationBlock>
</div>
```

6 Microstructure

6.1 Tokens (<w>)

6.1.1 Characterization

Most transcription conventions do not provide an exact and comprehensive definition of the unit *word*. Rather, they take the word definition of standard written orthography as a starting point and supplement this with rules for a selected number of special cases (e.g. abbreviations and spelling, words specific to spoken language such as interjections). A more precise definition should not, 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" (see ISO 24611). 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 referable 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 where the part of speech of the token can be encoded. Similarly, a **@lemma** attribute can be used to associate the token with a lemma, such as an uninflected dictionary entry form. Using a **@lemmaRef** attribute, a pointer to a definition of the lemma for the token, for example, in an online lexicon, can be provided.

An **@xml:lang** attribute can be used to record the language of an individual token. This can be necessary, for instance, when code-switches occur inside an utterance.

Since information encoded in **@type**, **@ana**, **@lemma** and **@xml:lang** attributes constitutes an annotation to the token, this kind of information can alternatively be recorded as a (free) annotation in a **** element (see 5.3). This is especially advisable if there is no 1:1 relationship between **<w>** elements and annotations on the lemma or part-of-speech level (see Example 9).

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 characterized as an accentuated syllable or lengthened sound by an appropriate **@type** attribute or, again, by referencing the **<seg>** element from a ****

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>** shall allow **<anchor>** as a child. Pauses inside tokens can occur and should be encoded as **<pause>** elements, as described in 6.2.

6.1.4 Examples

EXAMPLE 14 Use of <w> element

```
<!-- 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>

<!-- language encoded as attribute on the token -->
<u who="#SPK0" start="#T0" end="#T2">
 <w xml:id="w148" lemma="I" xml:lang="en">I</w>
 <w xml:id="w149" lemma="be" xml:lang="en">am</w>
 <w xml:id="w150" lemma="very" xml:lang="fr">enchanté</w>
 <w xml:id="w151" lemma="much" xml:lang="fr">mon</w>
 <w xml:id="w152" lemma="aware" xml:lang="fr">cher</w>
 <w xml:id="w153" lemma="of" xml:lang="fr">ami</w>
</u>

<!-- a token with an accentuated syllable -->
```

```
<!-- the accentuation being represented in a separate span element -->
<annotationBlock 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>
</annotationBlock>

<!-- 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 Characterization

Most transcription systems distinguish measured pauses and typed pauses, the latter being typically divided into a small number of types based on perceived length; they include "micro", "short", "medium" and "long". Pauses can occur outside speakers' utterances (see 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 made by 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 is 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 by means of 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

EXAMPLE 15 Use of <pause>

```
<!-- measured pause -->
<pause dur="PT1.2S"/>

<!-- typed pause -->
<pause type="micro"/>

<!-- typed pause with original form in a rend attribute-->
<pause type="micro" rend="(.)"/>

<!-- pause inside an utterance -->
<u who="#SPK0" start="#T0" end="#T2">
 <w>I</w>
 <w>am</w>
 <pause dur="PT1.2S"/>
 <w>aware</w>
 <w>of</w>
 <w>that</w>
</u>

<!-- 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 Characterization

Non-speech events comprise a very varied set of phenomena ranging from productions with an obvious communicative function (e.g. audible laughter or a visible shake of the head) and secondary modes of communication (e.g. body language, gestures and facial expressions) to events (e.g. "telephone rings") and activities (e.g. "rummages in pocket") that 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 to speech (laughter at the end of an utterance) vs. events simultaneous with speech (words uttered while laughing);

— events which can be attributed to a speaker ("cough", "nod", "laughter") vs. events which cannot ("telephone rings", "microphone topples over").

Most systems will at least contain instructions for audible events that are alternative to speech and which can be attributed to a speaker. Of such phenomena, the most frequently 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 real 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 (Chapter 8) provide three different elements for describing non-speech events:

— **<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, and 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) events that are not directly communicative events but which 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 **<annotationBlock>**, 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 **<annotationBlock>** 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 Examples

EXAMPLE 16 Use of <vocal> and <kinesic>

```
<!-- 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 annotationBlock -->
<!-- with start and end points -->
<annotationBlock 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>
</annotationBlock>

<!-- (backchannel) nodding as kinesic element on the level of annotationBlock -->
<!-- with speaker assignment and start and end points -->
<annotationBlock who="#SPK0" start="#T6" end="#T9">
   <!-- [...] -->
</annotationBlock>
<kinesic who="#SPK1" start="#T7" end="#T8">
    <desc>nods</desc>
</kinesic>
```

6.4 Punctuation (<pc>)

6.4.1 Characterization

Since spoken utterances rarely follow the grammar of the written standard, few transcription systems employ punctuation that follows 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). It is more common for the semantics of punctuation symbols to be redefined to match salient characteristics of spoken language. One common system based on prosody uses punctuation symbols to delimit intonation phrases and to characterize their final tone movement, and in the GAT system, for instance, a period marks the end of an intonation phrase with a low falling tone movement, the question mark to identify the end of a phrase with a high rising tone movement and so on. A similar system is used in DT. Other uses of punctuation symbols include the marking of repair sequences (e.g. a forward slash is used in HIAT for that purpose), 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 6.1.2) or a **@type** attribute on a **<seg>** element (for tone movements, see 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 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 that 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 not usually correspond directly to some event occurring in time, and 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

EXAMPLE 17 Use of <pc>

```
<!-- 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 Characterization

Most transcription systems have mechanisms to mark uncertainty in transcription (i.e. parts where the transcriber is not sure of what he/she has 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 **** element.

Completely incomprehensible passages should be represented by a **<gap>** element. The **@reason** attribute should then be given the value **incomprehensible**. A **@dur** attribute may be used to indicate the temporal duration of the passage. Alternatively or additionally, attributes from the **att.dimensions** class (e.g. **@unit** + **@quantity** or **@extent**) can 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; alternatively, it may occur on the same level as **<u>** or **<annotationBlock>** elements if the omission is of a more global nature. In the latter case, **@start** and **@end** attributes pointing to the timeline shall be provided.

6.5.4 Examples

EXAMPLE 18 Use of <unclear>, <choice> and <gap>

```
<!-- 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 Characterization

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 vary greatly between different conventions and is much 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, that is to say, 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 (e.g. "utterance" or "intonation phrase"). A **@subtype** attribute can be added to provide an additional subclassification (e.g. "declarative", "interrogative" for the mode of an utterance or "falling", "rising" for the final tone movement of an intonation phrase). An **@xml:id** attribute can be provided to make the entity addressable for stand-off 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 6.4) is often used to delimit and characterize these units.

6.6.4 Examples

EXAMPLE 19 Use of <seg> to divide <u>

```
<!-- 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 reason="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) -->
<-- final tone movement specified in a @subtype attribute -->
<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>
```

Annex A
(informative)

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>Beckhams talkshow interview</title>
         </titleStmt>

         <!-- ********************************** -->
         <!-- Distribution information, see 4.1.1 -->
         <!-- ********************************** -->
         <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>

         <!-- ********************************** -->
         <!-- Recording information, see 4.1.2 -->
         <!-- ********************************** -->
         <sourceDesc>
            <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 the -->
                  <!-- 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>
         </sourceDesc>
      </fileDesc>

      <profileDesc>

         <!-- ********************************** -->
         <!-- Participant information, see 4.2.1 -->
         <!-- ********************************** -->
         <particDesc>
```

```
            <person xml:id="SPK0" n="PAR" sex="1" role="interviewer">
               <persName>
                  <forename>Michael</forename>
                  <surname>Parkinson</surname>
                  <roleName>Sir</roleName>
               </persName>
               <age value="72"/>
               <birth when="1935-03-28"/>
            </person>
            <person xml:id="SPK1" n="VIC" sex="2" role="interviewee">
               <persName>
                  <forename>Victoria</forename>
                  <surname>Beckham</surname>
               </persName>
               <age value="33"/>
               <birth when="1974-04-14"/>
            </person>
            <person xml:id="SPK2" n="DAV" sex="1" role="interviewee">
               <persName>
                  <forename>David</forename>
                  <surname>Beckham</surname>
               </persName>
               <age value="32"/>
               <birth when="1975-05-02"/>
            </person>
         </particDesc>


         <!-- ****************************** -->
         <!-- Setting information, see 4.2.2 -->
         <!-- ****************************** -->
         <settingDesc>
            <place>
               <placeName>BBC studio London</placeName>
            </place>
            <setting>
               <activity>Talkshow host Michael Parkinson interviewing David and Victoria
                     Beckham about their relationship</activity>
            </setting>
         </settingDesc>

      </profileDesc>

      <!-- ******************************** -->
      <!-- Description of source, see 4.3 -->
      <!-- ******************************** -->
      <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">
            <desc>Orthographic transcription according to HIAT</desc>
         </transcriptionDesc>
      </encodingDesc>

      <revisionDesc>
         <change when="2015-04-27T10:16:06.469+02:00">Created by XSL transformation from
an
            EXMARaLDA basic transcription</change>
      </revisionDesc>
   </teiHeader>

   <!-- END TEI HEADER -->

   <text>
      <!-- ***************** -->
      <!-- Timeline, see 5.1 -->
      <!-- ***************** -->
```

```
<timeline unit="s">
   <when xml:id="T0"/>
   <when xml:id="T1" interval="2.18" since="#T0"/>
   <when xml:id="T2" interval="2.43" since="#T0"/>
   <when xml:id="T3" interval="2.70" since="#T0"/>
   <when xml:id="T4" interval="3.74" since="#T0"/>
   <when xml:id="T5" interval="4.71" since="#T0"/>
   <when xml:id="T6" interval="5.07" since="#T0"/>
   <when xml:id="T7" interval="7.58" since="#T0"/>
   <when xml:id="T8" interval="8.53" since="#T0"/>
   <when xml:id="T9" interval="11.36" since="#T0"/>
   <when xml:id="T10" interval="13.91" since="#T0"/>
   <when xml:id="T11" interval="15.47" since="#T0"/>
   <when xml:id="T12" interval="16.56" since="#T0"/>
   <when xml:id="T13" interval="17.85" since="#T0"/>
   <when xml:id="T14" interval="20.79" since="#T0"/>
   <when xml:id="T15" interval="21.32" since="#T0"/>
   <when xml:id="T16" interval="23.89" since="#T0"/>
   <when xml:id="T17" interval="29.19" since="#T0"/>
   <when xml:id="T18" interval="30.47" since="#T0"/>
   <when xml:id="T19" interval="31.90" since="#T0"/>
   <when xml:id="T20" interval="33.45" since="#T0"/>
   <when xml:id="T21" interval="35.67" since="#T0"/>
   <when xml:id="T22" interval="37.35" since="#T0"/>
   <when xml:id="T23" interval="38.42" since="#T0"/>
   <when xml:id="T24" interval="41.39" since="#T0"/>
   <when xml:id="T25" interval="42.35" since="#T0"/>
   <when xml:id="T26" interval="46.60" since="#T0"/>
   <when xml:id="T27" interval="47.50" since="#T0"/>
   <when xml:id="T28" interval="51.00" since="#T0"/>
   <when xml:id="T29" interval="52.15" since="#T0"/>
   <when xml:id="T30" interval="53.97" since="#T0"/>
   <when xml:id="T31" interval="56.28" since="#T0"/>
   <when xml:id="T32" interval="59.19" since="#T0"/>
   <when xml:id="T33" interval="59.79" since="#T0"/>
   <when xml:id="T34" interval="60.61" since="#T0"/>
   <when xml:id="T35" interval="61.36" since="#T0"/>
   <when xml:id="T36" interval="62.61" since="#T0"/>
   <when xml:id="T37" interval="63.13" since="#T0"/>
   <when xml:id="T38" interval="65.96" since="#T0"/>
</timeline>

<!-- ********************************** -->
<!-- The actual transcription, see 5 and 6 -->
<!-- ********************************** -->
<body>
   <!-- annotationBlock grouping u with dependent annotations, see 5.4 -->
   <annotationBlock who="#SPK0" start="#T0" end="#T9" xml:id="ag1">
      <!-- utterance, see 5.2 -->
      <u xml:id="u1">
         <!-- unit above the token and below the u level, see 6.6 -->
         <seg xml:id="seg0" type="utterance" subtype="declarative">
            <!-- (word) token, see 6.1 -->
            <w xml:id="w1">And</w>
            <w xml:id="w2">what</w>
            <w xml:id="w3">comes</w>
            <!-- uncertainty on the transcriber's part, see 6.5 -->
            <unclear>
               <choice>
                  <seg>
                     <w xml:id="w4">through</w>
                     <w xml:id="w5">is</w>
                  </seg>
                  <seg>
                     <w xml:id="w4a">to</w>
                     <w xml:id="w5a">as</w>
                  </seg>
               </choice>
            </unclear>
            <w xml:id="w6">your</w>
            <w xml:id="w7">determination</w>
```

```
      <!-- time information within a u element, see 5.2 -->
      <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"/>
      <!-- (measured) pause, see 6.2 -->
      <pause dur="PT0.3S"/>
      <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"/>
      <!-- typed (word) token, see 6.1 -->
      <w xml:id="w29" type="repetition">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>
      <!-- punctuation element, see 6.4 -->
      <pc xml:id="pc1">,</pc>
      <w xml:id="w44">don't</w>
      <w xml:id="w45">they</w>
   </seg>
</u>

<!-- annotation of tempo via reference to the timeline, see 5.3 -->
<spanGrp type="tempo">
   <span from="#T6" to="#T7">faster</span>
</spanGrp>

<!-- part-of-speech annotation via reference to token IDs, see 5.3 -->
<spanGrp type="pos">
   <span from="#w1" to="#w1">CONJ</span>
   <span from="#w2" to="#w2">RELPRO</span>
   <span from="#w3" to="#w3">V</span>
   <span from="#w4" to="#w4">ADV</span>
```