New work item proposal

Transcription of Multimodal Corpora

Potential experts

Thomas: please comment, complement.

- Bertrand Gaiffe
- Michael Kipp
- Han Sloetjes (MPI Nijmegen, developer of ELAN)
- Lou Burnard
- Kai Wörner (Hamburg, co-developer of EXMARaLDA)
- Christophe Parisse (MoDyCo / Paris Ouest)
- Helen Spencer-Oatey (BASE corpus, University of Warwick) or the TEI expert of her team (might be Lou...)?
- Ulrike Gut (Münster)?

Issues to be discussed

- Cf. <c> vs. <pc>
- Typology for segments => standardised?
- Meta-data for identifying originating tool
- Merging of annotation files - cf. document structure (multiple documents - TEICorpus)

Introduction

Context, multiple tools, multiple formats, multiple transcription conventions cf. Intro TS-jTEI
Complemetarity with specific annotation levels MAF, SynAF, SemAF
Joint ISO-TEI initiative (cf. MoU)

1 Scope

Export, import, interchange of multimodal resources
Transcribed + basic Partitur/multi-level annotations
TEI customization

2 Normative references

XML
TEI P5 (version?)

3 Terms and definitions

transcription
annotation
transcription convention
transcription tool

tier
…

4 General document structure

One, multiple documents? – TEICorpus?
primary - subordinate
tier
metadata

5 Metadata for the transcription of multimodal corpora

section defining original tool(s)?
section defining transcription convention(s) used? - important for calculating “appropriate” visualisations
mechanism for pointing to underlying audio/video file(s)? - not available in T5 so far
<particDesc> and <person> to define and describe speakers

```
<particDesc>
   <person xml:id="SPK0" sex="1">
     <persName>
       <abbr>DS</abbr>
     </persName>
     <!-- possibly further descriptive elements -->
   </person>
   <person xml:id="SPK1" sex="0">
     <persName>
       <abbr>FB</abbr>
     </persName>
   </person>
</particDesc>
```

6 Macrostructure

6.1 Characterisation in terms of annotation graphs

(see also transcription graphs in Schmidt 2005 and the discussion in Schmidt et al. 2009)

Annotation graphs as in Bird/Liberman (2001) plus:

- graph must be *fully anchored*
- nodes must be *fully ordered*
- arcs can (but need not) be assigned to one member of a set of *speakers*
- each arc must be assigned to exactly one member of a set of *categories*
- exactly one category is assigned to *type* T(ranscription), the remaining categories are assigned to either type A(nnotation) or D(escription)
- arcs assigned to categories of types T or A must also be assigned to a speaker
- partition of arcs:
    - all arcs assigned to the same speaker and the category of type 'T' → main tier for that speaker, constraint: no overlapping arcs (typically: orthographic transcription)
    - all arcs assigned to the same speaker and one of the categories of type 'A' → dependent tier(s) for that speaker, constraint: corresponding arcs in main tier of the same speaker (typically: linguistic annotation)
    - all arcs assigned to the same speaker and one of the categories of type 'D' → independent secondary tier(s) for that speaker (typically: descriptions of non-verbal behaviour)
    - a (maximally?) contiguous set of arcs in a main tier → segment chain → <u> element

6.2 Timeline (<timeline>)

@origin - the date and time when the recording starts? is not necessarily known with sufficient precision
contains ordered set of <when> elements
with obligatory @id
with optional @absolute attribute specifying the offset into the recording
values in @absolute must be strictly monotonic increasing

```
<timeline unit="s" origin="#T0">
  <when xml:id="T0" absolute="00:00:00"/>
  <when xml:id="T1" absolute="00:00:02.13"/>
  <when xml:id="T2" absolute="00:00:03.74"/>
  <when xml:id="T3" absolute="00:00:04.71"/>
  <when xml:id="T4"/>
  <when xml:id="T5" absolute="00:00:08.53"/>
  <when xml:id="T6" absolute="00:00:11.36"/>
  <when xml:id="T7" absolute="00:00:13.91"/>
  <when xml:id="T8" absolute="00:00:15.47"/>
  <!-- [...] more when elements -->
</timeline>
```

6.3 Utterances (<u>)

speaker attribution (@who)
temporal structure:

- either: obligatory @start and @end attributes (IDREFs to <when>),
- or: obligatory first and last anchor element with @synch attribute (IDREF to <when>)
- arbitrary numbers of <anchor> inside <u>,
- overlap represented implicitly through temporal structure (obligatory),
- other mechanisms (@trans='overlap') discouraged/forbidden

```
<u who="#SPK0">
  <anchor synch="#T1"/>Okay. <anchor synch="#T2"/>Très bien,
  <anchor synch="#T3"/>très bien. <anchor synch="#T4"/>
</u>
```

6.4 Dependent annotations (<spanGrp>)

using elements in a <spanGrp> (one spanGrp groups annotations of the same kind)
@from and @to can point to points in the timeline
@from and @to can also point to ids of other elements, but this type of reference can usually not be derived automatically from the annotation tool data model/format
see also the corresponding section in Romary & Witt (2012)
further annotation techniques (feature structures!) not precluded, but not in the scope of this document

```
<!-- additional annotations from a sup (=suprasegmentals) tier -->
<spanGrp type="sup">
   <span from="#T2" to="#T4">faster</span>
</spanGrp>
<!-- additional annotations from an en (=English translation) tier -->
<spanGrp type="en">
   <span from="#T1" to="#T2">Okay. </span>
   <span from="#T2" to="#T4">Very good, very good.</span>
</spanGrp>
```

6.5 Grouping of utterances and dependent annotations ( )

purpose: "local" annotated environments - each such or sequence of such s is a transcription in its own right
appropriate name/type for the div?

```
<div>
```

```
   <!-- the transcribed text from the primary tier -->
   <u who="#SPK0">
         <!-- [...] (see above) -->
   </u>
   <!-- additional annotations from a sup (=suprasegmentals) tier -->
    <spanGrp type="sup">
         <!-- [...] (see above) -->
    </spanGrp>
    <!-- additional annotations from an en (=English translation) tier -->
    <spanGrp type="en">
         <!-- [...] (see above) -->
    </spanGrp>
</div>
```

6.6 Independent elements outside utterances

typically speakerless <pause> and <incident> elements
also: non-verbal (<kinesic> etc.) possibly simultaneous to speech attributed to a speaker
must have @start and @end
encoded on the level of elements

```
<div>
     <!-- [...] u and spanGrp elements, see above -->
</div>
<!-- an incident from a nv (=nonverbal) tier describing nonverbal behaviour -->
<incident who="#SPK0" type="nv" start="#T3" end="#T6">
  <desc>right hand raised</desc>
</incident>
<div>
     <!-- [...] u and spanGrp elements, see above -->
</div>
```

7 Microstructure

7.1 Words

7.1.1 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.

Some transcription conventions have methods for representing an *assimilation*, i.e. a blending, of two or more words into one. Also common are methods for characterising a word as *incomplete* (cut-off or initialising a self repair sequence). Where certain syllables are pronounced more lengthened than in standard pronunciation, some transcription conventions mark this syllable accordingly.

7.1.2 Representation as <w>

=> specDesc
Attributes: @type=’assimilated’ on the later word for assimilated words, alternatively: @trans=’assimilated’ as in transitions between <u>, but the guidelines do not allow such an attribute here

7.1.3 Further constraints

since overlaps starting or ending inside a word occur, <w> must allow <anchor> as a child leading to mixed content
identification (@xml:id) is extremely helpful for feeding data into NLP tools, should be made obligatory

7.1.4 Examples

```
<!-- an utterance divided into words -->
<u>
<!-- [...] -->
   <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>

<!-- a word with a time anchor inside -->
<w xml:id="w152">a<anchor synch="#T3"/>ware</w>
```

7.2 Pauses

7.2.1 Characterisation

measured pauses vs. typed pauses (types: micro, short, medium, long)
pauses inside vs. pauses outside <u> (speaker attribution)

7.2.2 Representation as <pause>

7.2.3 Further constraints

Temporal plausability - length of measured pause should not contradict temporal information as encoded in timeline references

7.2.4 Examples

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

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

<!-- measured pause outside <u>, with its own start and end attributes -->
<pause dur="PT0.61S" start="TLI_10" end="TLI_11"/>
```

7.3 Audible non-speech events

7.3.1 Characterisation

breathing, laughing, coughing etc.
noises not attributable to a speaker (e.g. telephone rings)
visible non-speech events (e.g. nods)? → multimodal, not in the scope of this document?

7.3.2 Representation as <vocal>, <kinesic> or <incident>

7.3.3 Further constraints

order of contributions and elements on the same level:

1. ascending by position of @start in <timeline>
2. descending by position of @end in <timeline> (when @start are equal)
3. ascending by position of @who in <particDesc> (when @start and @end are equal)

7.3.4 Examples

```
<!-- coughing encoded as incident between words and anchors of a u -->
<u>
    <anchor synch="#T4"/>
    <w>dépend</w>
    <incident>
       <desc>cough</desc>
     </incident>
    <anchor synch="#T5"/>
    <w>un</w>
```

```
    <w>peu</w>
    <anchor synch="#T6"/>
  </seg>
</u>
```

7.4 Punctuation

7.4.1 Characterisation

punctuation according to orthography (e.g. period at the end of a grammatical sentence or comma introducing a subordinate clause in German) vs. puncutation representing properties of speech (e.g. comma representing a rising tone movement, slash representing the initialisation of a repair sequence) - the latter should ideally be mapped to a corresponding annotation element, but this is not always feasible
N.B.: In contrast to other elements, punctuation does usually not directly correspond to some event occuring in time → no start and end attribute possible

7.4.2 Representation as <c> (or <pc>?)

<pc> is probably more accurate (original proposal uses <c>)

7.4.3 Further constraints

7.4.4 Examples

```
<!-- punctuation represented as pc elements -->
<seg function="utterance">
  <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>.</pc>
</seg>
```

7.5 Uncertainty and incomprehensible passages

7.5.1 Characterisation

7.5.2 Representation as <unclear>

<unclear> with PCDATA represents uncertainty
alternatives should be each marked as unclear and grouped with a <choice> element
empty <unclear> elements for incomprehensible passages?

7.5.3 Further constraints

7.5.4 Examples

```
<!-- uncertain passage -->
<w>you</w>
<unclear>
  <w>should</w>
</unclear>
<w>let</w>

<!-- uncertain passage with alternatives -->
<w>you</w>
<choice>
 <unclear>
   <w>should</w>
 </unclear>
 <unclear>
   <w>could</w>
 </unclear>
</choice>
<w>let</w>
```

7.6 Units above the word and below the <u> level

7.6.1 Characterisation

Speaker's contributions can often be subdivided in chunks comprising more than one word and/ or pauses and/or non-audible speech events. These are the "sentence equivalents" of spoken language.
How these chunks are defined, distinguished and delimited varies greatly between different conventions (and is hotly debated). Two popular approaches: use pragmatic/syntactic criteria (--> notion of an utterance, e.g. HIAT/CHAT) vs. use prosodic critertia (--> notion of an Intonation phrases, e.g. GAT, DT)

7.6.2 Representation as <seg>

7.6.3 Further constraints

7.6.4 Examples

```
<!-- u divided into two seg elements -->
<div>
  <u who="#SPK0">
    <anchor synch="#T40"/>
    <seg function="utterance" type="declarative">
      <w xml:id="w319">And</w>
      <incident>
        <desc>unv.</desc>
      </incident>
      <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 function="utterance" type="interrogative">
      <incident>
        <desc>unv.</desc>
      </incident>
      <w xml:id="w325">you</w>
      <pc>,</pc>
      <w xml:id="w326">Victoria</w>
    </seg>
    <anchor synch="#T43"/>
  </u>
</div>
```

8 Bibliographical reference

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

Romary Laurent, Witt Andreas (2012). Data formats for phonological corpora. Handbook of Corpus Phonology Oxford University Press (Ed.) [http://hal.inria.fr/inria-00630289]

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

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] ; DOI : 10.4000/jtei.142

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

9 Annex

ODD spec

10 Annex - fully encoded example

11 Annex(es)

Mappings - macrostructure

12 Annex(es)

Mappings - microstructure