BS ISO 24624:2016

BSI Standards Publication

Language resource management — Transcription of spoken language

Documents
Document Sharing Hub

Full document details: BS ISO 24624:2016

National foreword

This British Standard is the UK implementation of ISO 24624:2016.

The UK participation in its preparation was entrusted to Technical Committee TS/1, Terminology.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

© The British Standards Institution 2016.
Published by BSI Standards Limited 2016

ISBN 978 0 580 84640 3

ICS 01.140.10

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 August 2016.

Amendments/corrigenda issued since publication

Date	Text affected

Documents
Document Sharing Hub

Full document details: BS ISO 24624:2016

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)

Documents
Document Sharing Hub
© ISO 2016

Full document details: BS ISO 24624:2016

COPYRIGHT PROTECTED DOCUMENT

© ISO 2016, Published in Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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

© ISO 2016 – All rights reserved

Documents
Document Sharing Hub

Full document details: BS ISO 24624:2016

Contents

Page

Foreword ... **v**

Introduction ... **vi**

1 Scope ... **1**

2 Normative references ... **1**

3 Terms and definitions ... **1**

4 Metadata ... **2**
- 4.1 Description of the electronic file (<fileDesc>) ... 2
  - 4.1.1 Distribution information (<publicationStmt>) ... 2
  - 4.1.2 Recording information (<recordingStmt>) ... 2
- 4.2 Description of circumstances (<profileDesc>) ... 4
  - 4.2.1 Participant information (<particDesc>) ... 4
  - 4.2.2 Setting information (<settingDesc>) ... 4
- 4.3 Description of source (<encodingDesc>) ... 5

5 Macrostructure ... **5**
- 5.1 Timeline (<timeline>) ... 5
- 5.2 Utterances (<u>) ... 6
- 5.3 Free dependent annotations (<spanGrp>,) ... 7
- 5.4 Grouping of utterances and dependent annotations (<annotationBlock>) ... 9
- 5.5 Independent elements outside utterances (<pause> and <incident>) ... 10
- 5.6 Inline paralinguistic annotation (<shift>) ... 10
- 5.7 Global divisions of a transcription ( ) ... 11

6 Microstructure ... **12**
- 6.1 Tokens (<w>) ... 12
  - 6.1.1 Characterization ... 12
  - 6.1.2 Representation as <w> ... 12
  - 6.1.3 Further constraints ... 13
  - 6.1.4 Examples ... 13
- 6.2 Pauses (<pause>) ... 14
  - 6.2.1 Characterization ... 14
  - 6.2.2 Representation as <pause> ... 14
  - 6.2.3 Further constraints ... 14
  - 6.2.4 Examples ... 15
- 6.3 Audible and visible non-speech events (<vocal>, <kinesic> and <incident>) ... 15
  - 6.3.1 Characterization ... 15
  - 6.3.2 Representation as <vocal>, <kinesic> or <incident> ... 16
  - 6.3.3 Examples ... 16
- 6.4 Punctuation (<pc>) ... 17
  - 6.4.1 Characterization ... 17
  - 6.4.2 Representation as <pc> ... 17
  - 6.4.3 Further constraints ... 17
  - 6.4.4 Examples ... 18
- 6.5 Uncertainty, alternatives, incomprehensible and omitted passages (<unclear>, <choice>, <gap>) ... 18
  - 6.5.1 Characterization ... 18
  - 6.5.2 Representation as <unclear> or <gap> ... 18
  - 6.5.3 Further constraints ... 18
  - 6.5.4 Examples ... 19
- 6.6 Units above the token and below the <u> level (<seg>) ... 20
  - 6.6.1 Characterization ... 20
  - 6.6.2 Representation as <seg> ... 20
  - 6.6.3 Further constraints ... 20
  - 6.6.4 Examples ... 20

AVA Documents
Document Sharing Hub

© ISO 2016 – All rights reserved

Full document details: BS ISO 24624:2016

Annex A (informative) **Fully encoded example** **22**
Annex B (informative) **Element and attribute index** **28**
Bibliography **31**

© ISO 2016 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 24624:2016

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*.

© ISO 2016 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24624:2016

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

© ISO 2016 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 24624:2016

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]

© ISO 2016 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24624:2016

Documents
Document Sharing Hub

INTERNATIONAL STANDARD ISO 24624:2016(E)

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

© ISO 2016 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24624:2016

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).

© ISO 2016 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 24624:2016

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
```

© ISO 2016 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24624:2016

```
            <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

© ISO 2016 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 24624:2016