

ISO 24624:2016-08 (E)

Language resource management - Transcription of spoken language

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