

CRITT process data repository metadata

The text below documents the metadata description of the process data in the CBS process data repository <http://copenhagen-dependency-treebank.googlecode.com/svn/trunk/translog/>.

Last updated: 25 July 2012

Study metadata (StudyMetaData.csv)

Study metadata is the top-level metadata element. It describes the experiment and the purpose of the experiment. It contains five categories of information:

ExperimentID is a unique identifier which is represented as a derived element in *Stimulus metadata* and *Recordings metadata*.

Abstract contains an abstract of the main study for which the process data have been collected.

Keywords lists the keywords of the experiment.

MainLiterature contains a reference to the main study for which data have been collected.

SecondaryLiterature contains references to other studies than the main study that have analysed data from the experiment.

Stimulus metadata (StimulusMetaData.csv)

Stimulus metadata describes static properties of the source texts that were used in the various experiments. It contains six categories of information:

StimulusID is a unique identifier which is represented as a derived element in *Recordings metadata*.

SourceLanguage states the language of the source text.

LengthWords states the number of words of the source text.

LengthCharacters states the number of characters of the source text.

Text contains the source text in its entirety.

Participant metadata (ParticipantMetaData.csv)

Participant metadata contains information about the participants from whom process data have been collected. It contains 21 categories of information:

ExperimentID is a derived identifier from *Study metadata* which links the participant explicitly to an experiment.

ExperimentParticipantID is a unique identifier which is represented as a derived element in *Recordings metadata*.

Sex of the participant.

YearOfBirth of the participant.

University that the participant attended.

Programme that the participant was enrolled into.

Student at the time of recording (yes/no).

DegreeStartedYear specifies the year in which the participant was enrolled into a university programme.

DegreeFinishedYear specifies the last year of the participant's university programme enrolment.

YearsTraining specifies the number of years the participant received translation specific instruction.

CertifiedTranslator specifies whether or not the participant has received formal authorisation to work as a translator and/or interpreter.

ExperienceYears specifies the number of years the participant has worked as a professional translator.

L1 of the participant.

L2 of the participant.

L3 of the participant.

OpticalAids specifies whether or not the participant uses optical aids such as glasses or contact lenses.

LeftEye specifies the diopetre for the left eye.

RightEye specifies the diopetre for the right eye.

EyeColour of the participant.

Recording metadata (RecordingMetaData.csv)

Recording metadata contains information about the all recordings in the repository. It contains 19 categories of information:

ExperimentID is a derived identifier from *Experiment metadata* which links the recording explicitly to an experiment.

StimulusID is a derived identifier from *Stimulus metadata* which links the recording explicitly to a text.

NameInStudy is the name given to the text in the main study for which the process data was collected.

NameInDatabase is the name given to the text in the SVN repository. NameInDatabase may be similar to StimulusID and/or NameInDatabase.

ExperimentType specifies the task that the participant carried out (e.g. translation, reading, copying, post-editing, etc.)

ExperimentTypeAdd provides additional information about the task.

FontType specifies the font type that the text was presented with for the specific recording.

FontSize specifies the font size that the text was presented with for the specific recording.

ExperimentParticipantID

UniqueParticipantID

LocationLink

RecordingData(MMDDYYYY) specifies the date, month and year for the recording.

EyeTrackerType specifies the eye tracking equipment that was used to collect the eye-tracking data.

RecordingSoftware specifies the eye tracking recording software that was used to collect the eye-tracking data.

EyeTrackerSoftwareVersion specifies the software version of the eye-tracking recording software.

Keylogger specifies the keylogging software that was used to collect the typing data.

KeyloggingSoftwareVersion specifies the software version of the keylogging software.

ExperimentalLocation specifies where the recording was carried out.

TargetLanguage specifies the language into which the source text was translated, copied, post-edited, etc.