

# Towards a standard “Emotion Language” for representing and annotating emotions

Working document of the W3C Emotion Incubator group

Date: 08 September 2006

Authors: Marc Schröder, ...

## Changelog

2006-09-08 Marc Schröder initial version based on use cases discussed in HUMAINE

### **1. Use cases and requirements**

The following use cases outline the types of scenarios in which the Emotion Language could be used.

#### **Use case 1: Annotation of emotional data**

##### ***Use case 1a: Annotation of plain text***

(i) Alexander is compiling a list of emotion words and wants to annotate, for each word or multi-word expression, the emotional connotation assigned to it. In view of automatic emotion classification of texts, he is primarily interested in annotating the valence of the emotion (positive vs. negative), but needs a “degree” value associated with the valence. In the future, he is hoping to use a more sophisticated model, so already now in addition to valence, he wants to annotate emotion categories (joy, sadness, surprise, ...), along with their intensities. However, given the fact that he is not a trained psychologist, he is uncertain which set of emotion categories to use.

(ii) ...

Requirements:

- scope of the emotion annotation: refer to one or several words of text
- emotion description:
  - emotion dimension: valence dimension, a scale from -1 to 1
  - emotion categories: set of labels, with intensity (from 0 to 1)
- other:
  - advice: which set of emotion categories should be used?

##### ***Use case 1b: Annotation of XML structures and files***

(i) Stephanie is using a multi-layer annotation scheme for corpora of dialog speech, using a stand-off annotation format. One XML document represents the chain of words as individual XML nodes; another groups them into sentences; a third document describes the syntactic structure; a fourth document groups sentences into dialog utterances; etc. Now she wants to add descriptions of the “emotions” that occur in the dialog utterances (although she is not certain that “emotion” is exactly the right word to describe what she thinks is happening in the dialogs): agreement, joint laughter, surprise, hesitations, indications of social power, ... These are emotion-related effects, but not emotions in the sense as found in the textbooks.

(ii) Paul has a collection of pictures showing faces with different expressions. These pictures were created by asking people to contract specific muscles. Now, rating tests are being carried out, in which subjects should indicate the emotion expressed in each face. Subjects can choose from a set of six emotion terms. For each subject, the emotion chosen for the corresponding image file must be saved into an annotation file in view of statistical analysis.

(iii) ...

Requirements:

- scope of the emotion annotation: refer to an XML node or a file
- emotion description:
  - find suitable name for the “emotion-related things” occurring in a dialog
  - emotion categories: open set of “emotion-related things” or closed set of six emotion labels

### ***Use case 1c: Chart annotation of time-varying signals (e.g., multi-modal data)***

(i) Jean-Claude wants to annotate audio-visual recordings of authentic emotional recordings. Looking at such data, he and his colleagues have come up with a proposal of what should be annotated in order to properly describe the complexity of emotionally expressive behaviour as observed in these clips. He is using a video annotation tool that allows him to annotate a clip using a “chart”, in which annotations can be made on a number of layers. Each annotation has a start and an end time.

The types of emotional properties that Jean-Claude and his colleagues want to annotate are many. They want to use emotion labels, but sometimes more than one emotion label seems appropriate – for example, when a sad event comes and goes within a joyful episode, or when someone is talking about a memory which makes them at the same time angry and desperate. Depending on the emotions involved, this co-occurrence of emotions may be interpretable as a “blend” of “similar” emotions, or as a “conflict” of “contradictory” emotions. The two emotions that are present may have different intensities, so that one of them can be identified as the major emotion and the other one as the minor emotion. Emotions may be communicated differently through different modalities (speech, facial expression, ...); it may be necessary to annotate these separately. Attempts to “regulate” the emotion and/or the emotional expression can occur: holding back tears, hiding anger, simulating joy instead... The extent to which such regulation is present may vary. In all these annotations, a given annotator may be confident to various degrees.

In addition to the description of emotion itself, Jean-Claude needs to annotate various other things: the object or cause of the emotion; the expressive behaviour which accompanies the emotion, and which may be the basis for the emotion annotation (smiling, high pitch, etc.); the social and situational context in which the emotion occurs, including the overall communicative goal of the person described; various properties of the person, such as gender, age, or personality; various properties of the annotator, such as name, gender, and level of expertise; and information about the technical settings, such as recording conditions or video quality. Even if most of these should probably not be part of an emotion annotation language, it may be desirable to propose a principled method for linking to such information.

(ii) ...

Requirements:

- scope of the emotion annotation: start and end times in a clip; sometimes also modality
- emotion description:
  - emotion categories, but potentially more than one; either with the same scope or as

overlapping stretches of time

- intensity of an emotion (for each emotion separately if several are present)
- regulation of an emotion, as a degree of simulation, masking, ...
- confidence of an annotation, for all aspects of the annotation
- other:
  - make connection with various other annotations, including meta-data about context and persons and annotation of behaviour

### ***Use case 1d: Trace annotation of time-varying signals (e.g., multi-modal data)***

(i) Cate wants to annotate the same clips as Jean-Claude, but using a different approach. Rather than building complex charts with start and end time, she is using a tool that traces some property scales, continuously over time. Examples for such properties are: the emotion dimensions arousal, valence or power; the overall intensity of (any) emotion, i.e. the presence or absence of emotionality; the degree of presence of certain appraisals such as intrinsic pleasantness, goal conduciveness or sense of control over the situation; the degree to which an emotion episode seems to be acted or genuine. The time curve of such annotations should be preserved.

(ii) ...

Requirements:

- scope of the emotion annotation: an entire clip
- emotion description: a sequence of samples describing the time course of various emotion-related properties:
  - emotion dimensions such as arousal, valence, or power, as scales from -1 to 1 or from 0 to 1
  - emotional intensity, as a scale from 0 to 1
  - emotional appraisals such as intrinsic pleasantness, goal conduciveness or sense of control, as scales from 0 to 1 or from -1 to 1
  - regulation: degree of simulation etc., as scales from 0 to 1

### **Use case 2: Automatic recognition / classification of emotions**

(i) Anton has built an emotion classifier from speech data which had been annotated in a way similar to use case 1b: emotion labels were assigned on a per-word basis, and the classifier was trained with the acoustical data corresponding to the respective word. Ten labels had been used by the annotators, but some of them occurred only very rarely. Based on a similarity metric, Anton merged his labels into a smaller number of classes. In one version, the classifier distinguishes four classes; in another version, only two classes are used. The classifier internally associates various probabilities to class membership. The classifier can either output only the one emotion that received the highest probability, or all emotions with their respective probabilities. Classifier results apply in the first step to a single word; in a second step, the results for a sentence can be computed by averaging over the words in the sentence.

(ii) ...

Requirements:

- scope of the emotion annotation: a stretch of time corresponding to word or sentence

boundaries as identified by a speech recognizer

- emotion description:
  - emotion categories, possibly several
  - probability for each emotion category
- other:
  - similarity metric for emotion categories
  - mapping mechanism for converting a larger set of emotion labels into a smaller set taking into account the similarity metric
  - mechanism for converting several word-based annotations into one sentence-based annotation

### **Use case 3: Generation of emotional system behavior**

#### ***Use case 3a: Affective reasoner***

(i) Ruth is using an affective reasoning engine in an interactive virtual simulation for children. Taking into account the current knowledge of the virtual situation, the affective reasoner deduces the appropriate emotional response. To do that, the situation is first analysed in terms of a set of abstractions from the concrete situation, capturing the emotional significance of the situation for the agent. These abstractions are called “emotion-eliciting conditions” or “appraisals” depending on the model used. These “appraisals” can then be interpreted in terms of emotions, e.g. emotion categories.

(ii) ...

Requirements:

- scope: a dialog or behavioral act
- emotion description:
  - a configuration of several “emotion-eliciting conditions” or “appraisals”, which may be gradual scales
  - an emotion category, possibly with intensity
- others:
  - a mapping mechanism analyzing the configuration of appraisals and generating the corresponding category

#### ***Use case 3b: Drive speech synthesis, facial expression and/or gestural behavior***

(i) Marc has written a speech synthesis system that takes a set of coordinates on the emotion dimensions arousal, valence and power and converts them into a set of acoustic changes in the synthesized speech, realized using diphone synthesis. If the speech synthesizer is part of a complex generation system where an emotion is created by an affective reasoner as in use case 3a, emotions must be mapped from a representation in terms of appraisals or categories onto a dimensional representation before they are handed to the speech synthesizer.

(ii) Catherine has built an ECA system that can realize emotions in terms of facial expressions and gestural behavior. It is based on emotion categories, but the set of categories for which facial expression definitions exist is smaller than the list of categories that are generated in use case 3a. A mapping mechanism is needed to convert the larger category set to a smaller set of approximately

adequate facial expressions.

Requirements:

- scope of the emotion annotation: full sentences / utterances
- emotion description:
  - emotion dimensions arousal, valence, control
  - emotion categories
- other:
  - mapping mechanism from appraisals or categories onto dimensions
  - mapping mechanism from larger to smaller set of categories

[this list to be extended during September 2006]

## ***2. Consolidated list of requirements***

[to be written as the result of a critical discussion of the use cases; October 2006?]

## ***3. Critical investigation of existing Emotion markup languages***

[discuss extent to which the HUMANE EARL and other languages meet these requirements]

[to be discussed once the requirements are consolidated; November 2006 – January 2007?]

## ***4. Options for a revised specification***

[depending on the outcome of the investigation of existing specs, propose and discuss options for a revised spec; February-June 2007?]