# The FML - APML language

*Maurizio Mancini*
*University of Paris 8*
*France*
*m.mancini@iut.univ-paris8.fr*

*Catherine Pelachaud*
*University of Paris 8, INRIA*
*France*
*catherine.pelachaud@inria.fr*

## 1 Introduction

In this paper we present a new version of the APML (Affective Presentation Markup Language, [**?**]) representation language, called FML-APML. This new version encompasses the tags of APML as well as other tags related, for example, to world references and emotional state. The presented language has been developed in the Greta framework [**?**]. Greta is an ECA (Embodied Conversational Agent) that starting from a representation of its *communicative intention*, plans the verbal (speech) and nonverbal signals (facial expressions, head movements, gestures) in order to convey it. We use the FML-APML language to model the agent's communicative intention.

## 2 Related work: APML

APML is an XML-based markup language for representing the agent's communicative intention and the text to be uttered by the agent [**?**]. APML tags refer to the possible information a person may seek to communicate: information on the world, on the speaker's mind and on the the speaker's identity. Based on the Poggi's work [**?**], the APML language encodes the first and second types of information in ECAs [**?**]. In APML, each tag corresponds to one of the communicative intentions described in [**?**], namely:

- *certainty*: this is used to specify the degree of certainty the agent intends to express.
  Possible values: *certain, uncertain, certainly_not, doubt.*
- *meta-cognitive*: this is used to communicate the source of the agent's beliefs.
  Possible values: *planning, thinking, remembering.*
- *performative*: this represents the agent's performative [**?**][**?**].
  Possible values: *implore, order, suggest, propose, warn, approve, praise, recognize, disagree, agree, criticize, accept, advice, confirm, incite, refuse, question, ask, inform, request, announce, beg, greet.*

- *theme/rheme*: these represent the topic/comment of conversation; that is, respectively, the part of the discourse which is already known or new for the conversation's participants.
- *belief-relation*: this corresponds to the *metadiscoursive* goal, that is, the goal of stating the relationship between different parts of the discourse; it can be used to indicate contradiction between two concepts or a cause-effect link.

  Possible values: *gen-spec, cause-effect, solutionhood, suggestion, modifier, justification, contrast.*
- *turnallocation*: this models the agent's metaconversational goals, that is, the agent's intention to take or give the conversation floor.

  Possible values: *take, give.*
- *affect*: this represents the agent's emotional state. Emotion labels are taken from the OCC model of emotion.

  Possible values: *anger, disgust, joy, distress, fear, sadness, surprise, embarrassment, happy-for, gloating, resentment, relief, jealousy, envy, sorry-for, hope, satisfaction, fear-confirmed, disappointment, pride, shame, reproach, liking, disliking, gratitude, gratification, remorse, love, hate.*
- *emphasis*: this is used to emphasize (that is, to convey its importance) what the agent communicates either vocally (by adding pitch accents to the synthesized agent's speech) or through body movements (by raising the eyebrows, producing beat gestures, etc.).

  Possible values: *low, medium, high.*

## 3 FML-APML overview

In the SAIBA framework [**?**][**?**], the FML language encodes the agent's communicative intentions. FML-APML is an evolution of APML and presents some similarities and differences. The FML-APML tags are an extension of the ones defined by APML, so all the communicative intentions that we can represent in APML are also present in FML-APML. We introduced the following changes in creating FML-APML:

- Temporization of tags: APML tags have a nesting structure imposed by the way in which the language is defined. For example the top-level tag must always be a `performative` tag. The other tags, for example the one representing the agent's `certainty`, must be nested inside a performative:

```
<apml>
     <performative type="inform">
          <rheme certainty="certain">
               I'm the Greta agent
          </rheme>
     </performative>
</apml>
```

  The timing of these tags (i.e. the starting and ending of a certain communicative intention) is inferred from the duration of text nested inside the tags. In the above example, the performative, affect and certainty communicative intentions have the same starting and duration time. It is not possible, for example, to extend the three communicative intentions for a time slightly longer than the spoken text.

  In FML-APML each tag contains explicit timing data, similarly to BML tags. We also maintain coherence between the two languages defined inside the SAIBA framework. So, in FML-APML we can freely define the starting and ending time of each tag, or make tags referring to each other using symbolic labels. This also allows us to specify tags that are not linked to any spoken text. That is, with FML-APML we can define the communicative intention of non-speaking agents: for example we can represent the listener's communicative intention (e.g. the listener can have the intention to communicate that it is approving what the speaker says).

- Emotional state: we have extended the way in which the agent's emotional state is coded. In the APML representation, we can only specify the actually expressed emotion. In FML-APML we can model more complex situations, for example, if the speaker is feeling a certain emotion but he hides it by showing another, fake, emotional state [**?**]. We base our extension on EARL [**?**].
- Information on the world: when communicating with others, we could have the intention of communicating some physical or abstract properties of objects, persons, events. For example, we can accompany speech with hand shapes that mimic the shape of an object, or perform large arm movements to give the idea of an "amazing" event. APML syntax allowed one to specify only some of these kinds of intentions, sometimes in a too generic way. In APML the signal information was erroneously considered instead of the communicative intention: for example, the `deictic` tag could be used to explicitly perform deictic gestures. In FML-APML, we can specify that the agent is referring to an entity in the world, and eventually one of its properties. We leave the behavior planning system with the task of deciding if, to refer to this entity, the agent has to perform a deictic gesture, mimic its property, etc.

In the next Sections we give an overview of the FML-APML syntax: we present the FML-APML tags; then we describe the tags' attributes and temporization.

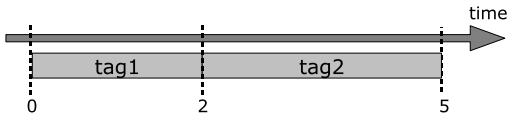
## 4 FML-APML tags: common attributes and synchronization

FML-APML tags are used to model the agent's communicative intention. Each tag represents a communicative intention (to inform about something, to refer to a place/object/person, to express an emotional state, etc.) that lasts from a certain starting time, for a certain number of seconds. The attributes common to all the FML-APML tags are:

- *name*: the name of the tag, representing the communicative intention modeled by the tag. For example, the name *performative* represents a performative communicative intention [**?**].
- *id*: a unique identifier associated to the tag; it allows one to refer to it in an unambiguous way.
- *type*: this attribute allows us to better specify the communicative meaning of the tag. For example, a performative tag has many possible values for the *type* attribute: *implore, order, suggest, propose, warn, approve, praise, etc.*. Depending on both the tag name (performative) and type (one of the above values), our Behavior Planning module determines the nonverbal behaviors the agent has to perform.
- *start*: starting time of the tag, in seconds. Can be absolute (time 0 corresponds to the start of the FML-APML file) or relative to another tag. It represents the point in time at which the communicative intention modeled by the tag begins.
- *end*: duration of the tag. Can be a numeric value (in seconds) relative to the beginning of the tag or a reference to the beginning or end of another tag (or a mathematical expression involving them). It represents the duration of the communicative intention modeled by the tag.
- *importance*: a value between 0 and 1 which determines the probability that the communicative intention encoded by the tag is communicated through nonverbal behavior. We describe this attribute in detail in Section **??**. It also modulates the number of modalities on which the communication happens, as explained in Section **??**.

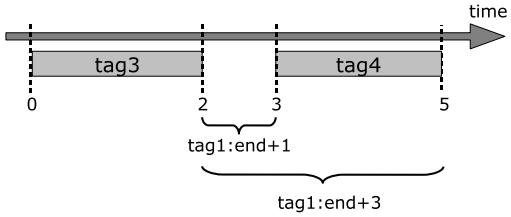
The timing attributes *start* and *end* also allow us to model the synchronization of the FML-APML tags. They both can assume absolute or relative values. In the first case, the attributes are numeric non-negative values, considering time 0 as the beginning of the FML-APML file. In the second case we can specify the starting or ending time of other tags, or a mathematical operation involving them. Note that the optional *end* attribute allows us to define communicative intentions that start at a certain point in time and last until new communicative intentions are defined. Here is an example of absolute and relative timings.

```
<FML-APML>
      <tag1 id="id1" start="0" end="2"/>
      <tag2 id="id2" start="2" end="3"/>
</FML-APML>
```




In the above FML-APML code, *tag*1 starts at time 0 and lasts 2 seconds; *tag*2 starts at time 2, and lasts 3 seconds.
All the timings are *absolute*, that is, they are both relative only to the beginning of the actual FML-AMPL file (equivalent to time 0).

```
<FML-APML>
      <tag3 id="id3" start="0" end="2"/>
      <tag4 id="id4" start="t1:end+1"
      end="t1:end+3"/>
</FML-APML>
```




In this case, the first tag is the same as before. On the other hand, *tag*2 has a *relative* timing as it starts as the first tag ends and lasts for 3 seconds.
FML-APML tags can be attached and synchronized to the text spoken by the agent. This is modeled by including a special tag, called *speech*, in the FML-APML syntax. Within this tag, we write the text to be spoken along with synchronization points (called *time markers*) which can be referred to by the other FML-APML tags in the same file. For example:

```
<FML-APML>
     <speech id="s1">
          <tm id="tm1"/>
               what are you
          <tm id="tm2"/>
               doing
          <tm id="tm3"/>
               here
          <tm id="tm4"/>
     </speech>

     <tag3 id="id3" start="s1:tm2" end="s1:tm4"/>
</FML-APML>
```

With the above code, we specify that the communicative intention of *tag*3 starts in correspondence with the word *doing* and ends at the end of the word *here*.

## 5 FML-APML importance attribute

We say that a message is important if it has a particular relevance to the Sender's goals: if a message is important we want to be sure that it is delivered to the receiver. The same situation occurs with communicative intentions.

Not all the communicative intentions we communicate to others have the same level of importance. Poggi et al. [**?**] note that, in the domain of goals (not necessarily communicative goals) different people may attribute a different importance to the same goal. For example, generous people attribute high importance to the goal of being helpful toward others; an independent person attributes high importance to the goal of making choices freely and without the others' help. De Carolis et al. [**?**] show that in nonverbal discourse planning the association of nonverbal signs to verbal information can be done by giving goals a *priority*.

The concept of *urgency* defined by Castelfranchi [**?**] seems to be related to importance: it is possible to sort the agent's goals depending on their urgency, and choose to display those goals which have a higher urgency value. Importance is also cited by Theune [**?**]. She claims that gesture frequency has to be increased if the speaker attaches a high importance to the message being communicated. For their conversational agents, Cassell et al. [**?**] choose to activate many modalities at the same time if the information importance is high. For example, information which is new or in contrast with respect to what has already been said is considered as having a higher priority and thus more modalities are activated. The importance of body actions is also referred to by Nayak et al. in [**?**]. In this case the importance level is directly translated into the priority of the corresponding body action, and higher priority actions are chosen first during behavior generation, while lower priority actions are discarded in case of conflict.

In FML-APML we introduce an attribute, common to each tag, called *importance*. Depending on its value, the agent may change the way the corresponding communicative intention is encoded. Similarly to the works proposed above, in our system the FML-APML *importance* attribute allows us to sort the concurrent agent's communicative intentions giving them a higher (resp. lower) priority if their importance is high (resp. low). We ensure that more important communicative intentions are communicated first by the agent while the least important intentions are eventually communicated by free communication modalities. Then, we use the same importance parameter to choose the multiplicity of multimodal behaviors. As the importance raises, we increment the number of modalities on which the agent's intentions are communicated. If, for example, importance is low and the agent is giving the user directions to reach a particular place in the environment, it produces only an iconic gesture. If importance is very high, it adds redundancy: the agent produces a deictic eye gesture (looking at the target in space), rotating the torso towards this position, while performing an iconic gesture.

## 6 Emotion tag

Emotion has a central role in communication and ECAs should be able to communicate their emotional state in order to increase effectiveness of interaction with humans. In the FML-APML language we have introduced the *emotion* tag, which models the speaker's *felt* and *expressed* emotional state. The former is the emotional state the speaker is really experiencing (which can be caused by an event, a person, a situation, etc.) while the latter is the one the speaker wants to communicate to the others. These two emotional states can be completely different: for example, a person can produce a "polite smile" to his superior even if he is angry at him. In general, people can *show* (expressed state is the felt one), *suppress* (the felt state is expressed the less possible), *mask* (the expressed state is different from the felt one) their emotional state [**?**]. In FML-APML we model these relations between felt and expressed emotional states by including the syntax of the EARL (Emotion Annotation and Representation Language) language, described in [**?**]. The emotion tag allows us to specify *complex emotional states*, as reported in [**?**]. We can for example model situations in which our agent is feeling a particular emotional state but simulates another emotion, hiding the felt one. This is done by controlling the felt and expressed emotional states with the *regulation* attribute of the emotion tag. The possible values of the regulation attribute are:

- *felt*: this indicates that the tag refers to a felt emotion;
- *fake*: this indicates that the tag refers to a fake emotion, an emotion that the agent aims at simulating;
- *inhibit*: the emotion in the tag is felt by the agent but it aims at inhibiting it as much as possible;

Let us consider the following example:

```
<FML-APML>
     <emotion id="e1" type="anger" regulation="felt" intensity="0.5" start="0" end="3"/>
     <emotion id="e2" type="joy" regulation="fake" intensity="0.9" start="0" end="3"/>
</FML-APML>
```

The agent's real emotional state is medium anger (the *regulation* attribute of the emotion tag is set to *felt*; intensity is 0.5, in a range going from 0 to 1) but it wants to hide it with an intense fake happiness (the *regulation* attribute of the emotion tag is set to *fake*; intensity is 0.9).

## 7 World tag

As explained in [**?**], while communicating with others, we seek to convey our knowledge about the world: objects and their characteristics (size, shape, location, etc.), events (real or abstract), places (relation, distance, etc.). Compared to APML, the FML-APML language introduces a *world* tag to indicate such kind of communicative intention. The tag has the following attributes:

- *ref_type*: the first attribute identifies the class of the referenced world entity: an object, a place, a time, an event. This attribute is required.
- *ref_id*: is an identifier that we can use to specify one or more world entities. This attribute is required.
- *prop_type* (optional feature): allows us to refer to a property of the referenced entity: its shape, location or duration.
- *prop_value* (optional feature): describes the value of the property specified with the previous attribute.

So, in FML-APML we can refer to an object in the world in a generic way, for example, if we want to refer to a book:

```
<FML-APML>
      <world id="w1" ref_type="object" ref_id="book"/>
</FML-APML>
```

Or, we can refer to the book which is on the table:

```
<FML-APML>
      <world id="w1" ref_type="object" ref_id="book" prop_type="location" prop_value="table"/>
</FML-APML>
```

## 8 Conclusions

In this paper we present FML-APML, a language which is used to model the communicative intention of an ECA. It is an extension of the previously developed APML language and it improves some of the APML weakness and missing features. We propose FML-APML as an implementation of the FML language of the SAIBA framework.