# The Next Step towards a Function Markup Language

Dirk Heylen[1], Stefan Kopp[2], Stacy C. Marsella[3],
Catherine Pelachaud[4], and Hannes Vilhjálmsson[5]

[1] Human Media Interaction, University of Twente
[2] Artificial Intelligence Group, Bielefeld University
[3] Information Science Institute, University of Southern California
[4] University of Paris 8, Inria
[5] Center for Analysis and Design of Intelligent Agents, Reykjavik University

**Abstract.** In order to enable collaboration and exchange of modules for generating multimodal communicative behaviours of robots and virtual agents, the SAIBA initiative envisions the definition of two representation languages. One of these is the Function Markup Language (FML). This language specifies the communicative intent behind an agent's behaviour. Currently, several research groups have contributed to the discussion on the definition of FML. The discussion reveals agreement on many points but it also points out important issues that need to be dealt with. This paper summarises the current state of affairs in thinking about FML.

## 1   Introduction

At the previous two installments of the Intelligent Virtual Agents conference, an update was presented about ongoing work on the Behavior Markup Language ([KKM⁺06, VCC⁺07]). BML is being discussed and specified by a representative number of research groups that strive for standardisation in the specification of behaviours of agents, to enable collaboration, sharing results and sharing actual working system components. BML is one of two representation languages in the SAIBA (Situation, Agent, Intention, Behavior, Animation) effort[1]. The other is FML, the Function Markup Language. FML will represent what an agent wants to achieve: its intentions, goals and plans.

At the first SAIBA workshop in Reykjavik in 2005 a preliminary proposal for FML was made ([VM05]) but since then research has focussed on BML rather than on FML. This year the first workshop on FML was proposed by the authors of this paper at AAMAS. The main aim of the workshop was to hear from the various researchers in the field about their wishes and suggestions regarding the functional specifications and to make an inventory of the issues that need to be resolved. This should enable one to lay down a roadmap for working out the next version of FML. In this paper we will review the current state of affairs

---

[1] http://wiki.mindmakers.org/projects:saiba:main/

in the discussion, representing and synthesizing the various views that emerged from the contributions and presenting the central issues that need to be tackled in the next stage of development. First we present some specific proposals for parts of FML from a selection of papers to provide an idea of the range of topics subsumed under it and of the range of variation between authors. Next, we take a broader view and look at the more general issues that arose.

## 2  Towards an Inventory of FML Tags

The contributions to the AAMAS workshop were of various types. Some raised general issues and others made specific proposals for (parts of) an FML specification. Several papers combined both aspects. In this section we will look at suggestions that were made for the attributes that FML might need to specify. First we consider the kinds of dimensions that were suggested in the discussion, often based on existing systems (hence the title legacy for the next section) and next we look at some specific proposals for attributes along the various dimensions.

### 2.1  Legacy

Vilhjálmsson and Thórisson ([VT08]), provide a brief history of function representation. Among the proposals for the representation of functions are the tags used in the BEAT system and the first proposal for FML within SAIBA. Several other contributions to the discussion list FML-related or FML-inspired information in current systems. To get an idea of the range of attributes that are involved and a sense of the overlap between proposals we briefly present five suggestions.

*BEAT/Spark.* The terms FML and BML were first employed to describe the tags set used in the BEAT system [CVB01] as part of Spark [Vil05]. In this case FML was a mark-up language for texts describing several discourse phenomena related to content and information structure (theme/rheme, emphasis, contrast, topic-shifts) and interaction processes (turn-taking and grounding).

*FML2005.* The FML break-out group at the 2005 SAIBA workshop in Reykjavik proposed to divide FML tags into a set of basic functional or semantic *units* and a set of *operations*. The *units* are typical elements present or occuring during a communicative event. The initial list comprised the following units.

 – Participants
 – Turns
 – Topic
 – Performative (speech act)
 – Content (proposition)

The operations that were suggested comprised:

 – Emphasis
 – Contrast

- Illustration
- Affect
- Social (or relational) goals
- Cognitive operations (e.g. difficulty of processing) and certainty

FML2005 shows a large overlap with the earlier proposal. An important difference is that it goes beyond the linguistic and conversational aspects and also includes elements of the mental state. Another important difference is that by introducing abstract units of discourse, FML2005 does not assume the annotation of an already produced text.

*TLCTS.* The paper by [SVJ08] on the Tactical Language and Culture Training System (TLCTS) considers intent planning mainly as a decision as to which communicative act to perform. This is (usually) a speech act. The communicative act specifies (1) the nature of the act (request, accept proposal...), (2) its modulation (politiness level, force), and (3) encoded contextual parameters (directed at male or female, role and status...). In order to be able to generate the appropriate behaviour in context, knowledge about the dialog, the world and the target culture is assumed to be needed.

Inspired by Traum and Hinkelman's typology of speech acts ([TH92]) the TLCTS system incorporates the following functions of communicative acts.

- Turn-taking: take-turn, release turn, keep-turn, assign-turn
- Grounding: initiate, continue, acknowledge, repair, request-repair, request-acknowledgement, cancel
- Core speech acts: inform, wh-question, yes-no question, accept, request, reject, suggest, evaluate, request-permission, offer, promise
- Argumentation: elaborate summarize, clarify, q&a, convince, find-plan

The overlap with the previous proposals on the main dimensions is again obvious. Turn-taking, grounding and speech acts are the main aspects that are considered. Interesting is the attention that is paid to contextual elements and the proposal that a third language called Context Markup Language should be devoted to them.

*APML.* The Greta framework [Pel05] uses APML as a mark-up language to encode the communicative intentions of an agent. The tags provide information about the following dimensions.

- The degree of certainty
- The meta-cognitive source of information (thinking, remembering, planning)
- The speech act (called performative),
- Information structure of the utterance (theme/rheme)
- Rhetorical relations such as contradiction or cause-effect (called belief-relations by the authors)
- Turn allocation
- Affect
- Emphasis

These tags are based on the work of Isabella Poggi [Pog07]. In [MP08], Mancini and Pelachaud propose an extention of APML, FML-APML, which is different in the following respects. The timing attributes of the tags are made consistent with the suggestions made for BML ([KKM+06, VCC+07]). This also makes it possible to specify the communicative intentions of non-speaking agents. The deictic tags were changed to remove elements that properly belong to BML. Furthermore, the emotional state tags are made more complex to allow for faked emotions (based on EARL [SPL06]). Also added is a tag to describe the importance of an intention.

*Virtual Human Project.* The nonverbal generation module of the virtual human architecture developed at ICT [LDMT08], contains several FML-inspired concepts, among which are the communicative intent (speech acts), cognitive operators that drive the gaze state of the virtual humans and elements that relate to emotional states and coping processes. The gaze model associates behaviours with what are called *cognitive operators* by providing a specification of the form and function of gaze patterns. These functions specify detailed reasons behind a particular gaze behaviour related to a combination of four categories of determinants: conversation regulation, updating of an internal cognitive state (desire, intention...), monitoring for events and goal status and coping strategy.

Although there are clear differences in the number and kind of dimensions that the various contributions to the AAMAS workshop consider for inclusion in FML, overall there is a big overlap between the various proposals. Prominent recurrent dimensions are *dialogue act* (or speech act), *turn-taking*, *grounding* actions, *content* (propositional content, discourse relations), *information structure* (emphasis, given/new, theme/rheme), *emotion, affect* and *interpersonal or social* relations.

We now turn to the kinds of attributes and values that are being proposed for the various dimensions. Some of the dimensions that have been proposed in the discussion but that are not included in this list will be introduced below as well.

## 2.2   Tags Proposed

When it comes to specifying the attributes of the various dimensions in more detail, it is to be expected that different views on them emerge, depending on the detail of specification that one wants to achieve, the complexity of the system that one is building or the specific demands of the application domain or the theoretical stance one takes. In the following paragraphs we provide a sketch of the elements that are proposed for the various dimensions and the variation between different research groups. The first dimension is an aspect of the context that was already present in FML2005 and worked out in more detail by [KS08]: participant information. Other dimensions discussed below are *communicative actions* (including turn-taking, grounding and speech acts), *content, mental state* and *social-relational goals*.

*Person characteristics.* Krenn and Sieber [KS08] provide a tentative list of characteristics of the communication partners that may need to be represented organised along three dimensions: person information, social aspects and personality and emotion. Here we describe the person characteristics only. The other elements such as emotion, personality and social aspects are covered below under separate headings. Person information, according to Krenn and Sieber should include the following information: an identifier, name, gender, type (human/agent), appearance, and voice. Also in the original SAIBA proposal on FML, information on participants was included. Among the features associated with it were role and id.

*Communicative Actions.* The main types of actions that are considered in the various proposals for FML are turn-taking actions, the actions involved in grounding a conversational action and a specification of the speech act as such (also called dialogue act or performative). The following table presents a few proposals for categorizing the turn-taking variables.

<div align="center">

*Turn-taking*

</div>

[KPL08] take-turn, want-turn, yield-turn, give-turn, keep-turn
[MP08]   take floor, give floor
[SVJ08]  take-turn, release-turn, keep-turn, assign-turn

For the process of grounding, Samtani et al. ([SVJ08]) propose the following actions (based on [TH92]).

<div align="center">

*Grounding*

</div>

[SVJ08] initiate, continue, ack, repair, req-repair, req-ack, cancel

As one might expect, the lists of performatives in the various proposals differ more widely than, for instance, the lists for turn-taking. This may be due to the fact that there is no agreement within linguistics but also because different applications may need more or less specific acts to be encoded. Compare, for instance, [MP08] with [SVJ08].

<div align="center">

*Speech Act*

</div>

[MP08]   implore, order, suggest, propose, warn, approve, praise, recognize, disagree, agree, criticize, accept, advice, confirm, incite, refuse, question, ask, inform, request, announce, beg, greet
[SVJ08]  inform, whq, ynq, accept, request, reject, suggest, eval, req-perm, offer, promise

Samtani et al. ([SVJ08]) also use acts that are involved in argumentation processes (see above) which might be put under the heading of speech act and others that are more like rhetorical relations.

*Content.* With respect to content, most papers do not go into the specific formalism to represent propositional content, assuming that some formal (logical) language will be used. Besides specifying propositional content itself, one also needs to take into account how it is organized.

On the level of a sentence, information structure is concerned with emphasis, given and new information or the related notions of theme and rheme. On a

discourse level, organization in topics is important as are rhetorical relations that hold between different parts of the discourse. This point is less settled. [SVJ08], for instance use items such as *elaborate* and *summarize* or *clarify* as part of the argumentation acts, which are typically rhetorical relations, besides actions such as *convince* or *find-plan*. [MP08] use the term *belief-relation* for relations between different parts of the discourse, including *gen-spec*, *cause-effect*, *solutionhood*, *suggestion*, *modifier*, *justification* and *contrast*.

*Mental State.* Most contributions to the FML discussion agree that some sort of representation of the emotional state of an agent is needed as emotions clearly contribute to the motivation of a communicative act. Several authors ([LDMT08] and [MP08], for instance) argue for the possibility to represent emotions in multiple ways within FML. Besides multiple representations for emotions, authors argue also for allowing distinct specifications of felt and expressed emotional states ([MP08], [KS08]), or leaked and felt emotions ([LDMT08]).

Besides emotional states, other mental states and processes that several authors feel should be represented in FML are cognitive processes such as *planning*, *thinking* and *remembering* ([MP08]). To account for gaze behaviours, Lee et al. [LDMT08] propose very detailed cognitive operations that include monitoring events and goal status, planning, conversation regulation functions and elements of coping.

*Social-Relational goals.* In the various proposals for elements to include in FML, one of the recurrent type of variables relates to the social psychological domain, i.e. the interpersonal variables that play a role in shaping interaction. So far no coherent picture of how to treat this has emerged, though several contributions to the FML discussion provide suggestions. Bickmore ([Bic08]), for instance, introduces relational (interpersonal) stance functions. His specific proposal was made for the context of agents involved in health-counseling. Everyone seems to agree that the social-relational dimension is an important element to take into account in FML, but this will need further elaboration.

Besides further elaboration of the various dimensions, work on the specification of FML is also faced with a number of general issues. We present several of these in the next section.

## 3   Issues

The SAIBA framework distinguishes three processes in the generation of multimodal communicative actions: intention planning, behavior planning, and behavior realization. FML is supposed to be concerned with specifying the intentions of an agent whereas BML is one kind of specification of the behaviour that results. A first question that arises is whether specifications of intentions and behaviours are sufficient for the multi-modal communicative action generation or whether this simple distinction misses out on aspects that do not fit in either of these categories. A second question involves the notion of intention itself. Finally, an important issue is the architecture of a complete generation system and

how FML and BML specifications fit in. We discuss each of these in turn. With respect to the first question, an important topic that appeared to be a concern of several authors is that of context.

## 3.1   The Role of Context

An issue that appears in various forms in several contributions to the FML discussion ([Bic08], [KS08], [SVJ08] and [Rut08] in particular) relates to the question of how contextual parameters that influence behaviours in a conversation should be treated. Should contextual parameters be part of FML or should they be covered by some other module? Samtani et al. ([SVJ08]) provide the example of greetings, the form of which depends on the time of day "Buenas Dias", "Buenas Tardes", "Buenas Noches". This is part of the *environmental context*. Others context variables are part of the dialog context including "the history of interactions", the "topics discussed" and the "cultural context".

Bickmore [Bic08] considers a particular form of context, for which he uses the term *frame* introduced by Bateson ([Bat54]) and also used by Goffman in ([Gof74]). People act differently in contexts that are differently framed and the same sentence uttered in a different frame may be associated with completely different intentions (said jokingly or not, for instance). Bickmore therefore proposes to add contextualisation tags to FML. The kinds of tags he uses currently for counseling agents in counselor-patient interactions are: *task* (information exchange), *social* (social chat, small talk), *empathy* (comforting interactions) and *encourage* (coaching, motivating, cheering up).

An issue to be solved is whether these context variables should be part of the Function Markup Language or whether they should be treated in some other way. The idea of a Contextual Markup Language featured in the AAMAS workshop as one way to frame the problem.

## 3.2   Intentions

Within the SAIBA framework, FML is described as the language that specifies the communicative intent behind an agent's behaviour. As the examples above show, not all of what the contributors to the workshop propose can be called intentional. The original conception of "intentions" in the SAIBA framework may have been derived from its predominance in speech act theory ([Aus62], [Sea69]). This deals mainly, if not exclusively, with intentions as determinants of communicative actions, as is highlighted by Grice's definition of non-natural meaning [Gri75].

> $S$ nonnaturally means something by an utterance $x$ if $S$ intends $(i_1)$ to produce by uttering $x$ a certain response $(r)$ in an audience $A$ and intends $(i_2)$ that $A$ shall recognize $S$'s intention $(i_1)$ and intends $(i_3,)$ that this recognition on the part of $A$ of $S$'s intention $(i_1,)$ shall function as $A$'s reason, or a part of his reason, for his response $r$.

Clearly, in the prototypical case of communication these intentions and the recognition of intentions play an important role, but the question is whether they are

the sole determinants of communicative behaviours and if not whether the other determinants need not be taken into account in FML as well. This will depend on one's point of view. Contextual parameters already constitute one set of determinants that may not fit well with the strict idea of intentions. In this case, there is a growing consensus that these should be separated out.

But there are other determinants of behaviour that are not intentions that one might want to include in FML. Several contributions to the discussion on FML suggest to include emotion tags into FML. Clearly, affective parameters will be an important determinant in the behaviours displayed by robots and virtual humans, however, what one is trying to model may be phenomena that escape conscious control and thus the term intention in inappropriate. This is epitomized in the case of leakage, where the behaviour does not communicate the emotion in the "nonnatural" way as above but as a symptom or "natural" meaning, for instance indicated by the trembling of the voice in case of nervousness. In the general framework one needs to take into account that there are several semiotic and intentional processes operating in communicative settings ([HtM08], [LDMT08]). It should be noted that the original name FML (Function Markup Language) implies a description of communicative *function*, a term that covers more than conscious intent.

## 3.3   Communicative Actions

Natural language utterances and nonverbal communicative acts are the most prominent kinds of communicative actions one is inclined to think of. However, as several authors have pointed out ([Gof76], for instance), communicative actions can also be completely extra-linguistic or certain non-linguistic actions can perform certain communicative functions. A typical case is that of contributions to grounding by performing the action that was previously requested showing that one has understood and accepted the request.

Kopp and Pfeiffer Leßman ([KPL08]) discuss a scenario of interactions between a human and the virtual agent Max where they collaborate in assembling Baufix parts. Here also, dialogue moves are heavily interwined with manipulative actions and both are considered forms of what the authors call *interaction moves* as a generalisation of dialogue moves.

Other prototypical cases where nonlinguistic actions become communicative occur when one performs an action ostentatiously. One can thereby communicate to an observer that one intends to perform the action and that one wants the observer to recognize the intention. This point is related to the previous one, in that many of such cases can be considered communicative acts by means of a specific semiotic process called demonstration ([Cla96]).

The general question for FML is thus: what to consider a communicative act and how to related communicative and noncommunicative acts.

## 3.4   Architecture

The SAIBA framework assumes a three step process of intention planning, behavior planning and behavior realization. In this model, it is relatively easy to

define the functionality of the FML and BML languages. However, as several authors have pointed out, the situation becomes less clear when one looks at the generation steps in detail and in particular at existing modules.

Lee et al. ([LDMT08] look at the particular case of natural language generation in which the reference architecture that is commonly assumed ([RD00]) defines three steps: document planning, microplanning and realization. The first part can – grosso modo – be interpreted as a form of intent planning, whereas the other two belong to the behaviour planning and realisation stage in SAIBA. But the question is whether FML representations will be compatible with existing generation systems.

A related point is made by Ruttkay ([Rut08] and by Krenn and Sieber ([KS08]) who write:

> This [...] brings us to another crucial aspect for the design of representation languages, i.e., the processing components used in ECA systems. We need to study which subsystems are implemented, what are the bits and pieces of information that are required as input to the individual processing components, and what kinds of information do the components produce as output. Especially if we aim at developing representations that will be shared within the community, there must be core processing components that are made available to and can be used by the community.

The first lesson of this for the development of FML and BML is thus that it is not enough to think about what should be in the language by looking at how conversations work, for instance, but also look into the modules that should make use of the languages, in particular modules that are currently already implemented. The second, might be that together with the specification of a language one needs to develop modules that can use them. This is certainly a major challenge for the definition of a Function Markup Language.

## 4   Conclusion

Although the work on jointly specifying a Function Markup Language has just started, the first results are promising in several respects. There seems to exist a rather high degree of consensus on the kinds of dimensions that need to be represented in FML. On the other hand, however, there are several dimensions for which the proposals on how to shape them diverge widely. Together with the general issues defined in the previous section on context, on definitional issues and on the embedding of FML in working modules, this means that the definition of a commonly agreed upon specification language for communicative intent still requires a lot of discussion in the wider ECA community. The current forum for discussion can be found at http://wiki.mindmakers.org/projects:fml:main/.

# References

[Aus62]     Austin, J.A.: How to Do Things with Words. Oxford University Press, London (1962)

[Bat54]     Bateson, G.: A theory of play and fantasy. Steps to an ecology of mind. Ballantine, New York (1954)

[Bic08]     Bickmore: Framing and interpersonal stance in relational agents. In: Why Conversational Agents do what they do. Functional Representations for Generating Conversational Agent Behavior, Estoril. AAMAS 2008 (2008)

[Cla96]     Clark, H.H.: Using Language. Cambridge University Press, Cambridge (1996)

[CVB01]     Cassell, J., Vilhjálmsson, H., Bickmore, T.: BEAT: the behavior expression animation toolkit. In: Proceedings of ACM Siggraph, Los Angeles, pp. 477–486 (2001)

[Gof74]     Goffman, E.: Frame Analysis. Penguin, Harmondsworth (1974)

[Gof76]     Goffman, E.: Replies and responses. Language in Society 5(3), 2257–2313 (1976)

[Gri75]     Grice, H.P.: Meaning. The Philosophical Review 66(3), 377–388 (1975)

[HtM08]     Heylen, D., ter Maat, M.: A linguistic view on functional markup languages. In: Why Conversational Agents do what they do, Estoril. Functional Representations for Generating Conversational Agent Behavior. AAMAS 2008 (2008)

[KKM+06]    Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thórisson, K., Vilhjálmsson, H.: Towards a common framework for multimodal generation in ECAs: the behavior markup language. In: Gratch, J. (ed.) Intelligent Virtual Agents 2006, pp. 205–217. Springer, Heidelberg (2006)

[KPL08]     Kopp, S., Pfeiffer-Leßmann, N.: Functions of speaking and acting. In: Why Conversational Agents do what they do. Functional Representations for Generating Conversational Agent Behavior. AAMAS 2008, Estoril (2008)

[KS08]      Krenn, B., Sieber, G.: Functional mark-up for behaviour planning. theory and practice. In: Why Conversational Agents do what they do. Functional Representations for Generating Conversational Agent Behavior. AAMAS 2008, Estoril (2008)

[LDMT08]    Lee, J., De Vault, D., Marsella, S., Traum, D.: Thoughts on FML: Behavior generation in the virtual human communication architecture. In: Why Conversational Agents do what they do. Functional Representations for Generating Conversational Agent Behavior. AAMAS 2008, Estoril (2008)

[MP08]      Mancini, M., Pelachaud, C.: The FML-APML language. In: Why Conversational Agents do what they do. Functional Representations for Generating Conversational Agent Behavior. AAMAS 2008, Estoril (2008)

[Pel05]     Pelachaud, C.: Multimodal expressive embodied conversational agents. In: Multimedia 2005, pp. 683–689. Springer, New York (2005)

[Pog07]     Poggi, I.: Mind, hands, face and body. A goal and belief view of multimodal communication. Weidler, Berlin (2007)

[RD00]      Reiter, E., Dale, R.: Building Natural Language Generation Systems. Cambridge University Press, Cambridge (2000)

[Rut08]     Ruttkay, Z.: Situation and agency in the SAIBA framework, and consequences for FML. In: Why Conversational Agents do what they do. Functional Representations for Generating Conversational Agent Behavior. AAMAS 2008, Estoril (2008)

[Sea69]     Searle, J.R.: Speech acts: An essay in the philosophy of language. Cambridge University Press, Cambridge (1969)

[SPL06]     Schröder, M., Piker, H., Lamolle, M.: First suggestions for an emotion annotation and representation language. In: International Conference on Language Resources and Evaluation: Workshop on Corpora for Research on Emotion and Affect, Genova, Italy, pp. 88–92 (2006)

[SVJ08]     Samtani, P., Valente, A., Johnson, W.L.: Applying the SAIBA framework to the tactical language and culture training system. In: Why Conversational Agents do what they do. Functional Representations for Generating Conversational Agent Behavior. AAMAS 2008, Estoril (2008)

[TH92]      Traum, D.R., Hinkelman, E.A.: Conversation acts in task-oriented spoken dialogue. Computational Intelligence 8, 575–599 (1992)

[VCC+07]    Vilhjálmsson, H., Cantelmo, N., Cassell, J., Chafai, N.E., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A.N., Pelachaud, C., Ruttkay, Z.M., Thórisson, K., van Welbergen, H., van der Werf, R.J.: The behavior markup language: Recent developments and challenges. In: Pelachaud, C., Martin, J.-C., Andre, E., Collet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 90–111. Springer, Heidelberg (2007)

[Vil05]     Vilhjámsson, H.: Augmenting online conversation through mediated discourse tagging. In: Proceedings of the 6th Annual Minitrack on Persistent Conversation. Hawaii International Conference on System Sciences, Hawaii. IEEE, Los Alamitos (2005)

[VM05]      Vilhjálmsson, H., Marsella, S.: Social performance framework. In: Workshop on Modular Construction of Human-Like Intelligence at the 20th National AAAI Conference on Artificial Intelligence, Pittsburgh, PA (2005)

[VT08]      Vilhjálmsson, H., Thórisson, K.: A brief history of function representation from gandalf to saiba. In: Why Conversational Agents do what they do. Functional Representations for Generating Conversational Agent Behavior. AAMAS 2008, Estoril (2008)