Functional Mark-up for Behaviour Planning: Theory and Practice

Brigitte Krenn[+±], Gregor Sieber[+]
[+]**Austrian Research Institute for Artificial Intelligence**
Freyung 6, 1010 Vienna, Austria

[±]**Research Studio Smart Agent Technologies**
Hasnerstrasse 123, 1160 Vienna, Austria

1. Introduction

We approach the discussion of requirements for an FML from a high-level perspective on communication and the current state of developments in ECA communication. From a general point of view questions arise such as: Who is communicating to whom in which socio-cultural and situational context. What is the overall interaction history of the communication partners, and what is the history of the ongoing dialogue. What is the intention of the communication and what is its content. Transferring these questions to the ECA domain, at least leads to questions of modelling the virtual character's persona including some notion of personality and emotion, and of modelling the communication act itself, be it in terms of real-time action and response or in terms of generating a complete dialogue scene in one go.

Our goal is mainly to come up with open questions and core topics regarding a possible scope of an FML given the current state of art in ECA communication. From a practical point of view, we start from a narrowed down perspective on modelling the communication partners and the communication act.

In section 2, we give a brief outline of the current state of ECA development and its implications for the creation of a commonly used mark-up or representation language at the interface of intent and behaviour planning. We propose a set of person characteristics and aspects of communication acts that need to be considered in the specification of a functional mark-up language. This is followed by a discussion of some basic building blocks relevant for the computation of communicative events (section 3). In section 4, we finally point out that one of the main challenges of FML lies in finding a trade-off between detailed semantic descriptions and interoperability of system components. We round up our considerations with some words of caution regarding the feasibility and desirability of a clear-cut separation between intent and behaviour planning.

2. Current Situation in ECA Development -- Implications for the Creation of a Functional Mark-up Language FML

Work on computational modelling of communicative behaviour is tightly coupled with the development of Embodied Conversational Characters (ECAs). In ECA systems, communicative events consist of (i) face-to-face dialogues between an interface character and a user [Matheson et al., 2003], (ii) an interface character presenting something to the user [Nijholt, 2006], (iii) two or more characters communicating with each other in a virtual or mixed environment, e.g. [Rehm & Andre, 2005]. On the one hand, there are ECA systems where only the generation side of multimodal communicative behaviour is simulated as it is the case with presenter agents where the whole dialogue scene is generated in one go, e.g. the NECA system [Krenn, 2003]. On the other hand, there are systems where the whole action-

reaction loop of communication is computed, i.e., the system interprets the input of a communication partner and then generates the reactions of the other communication partner(s) and so forth. See the REA system [Cassell et al., 1999] as an early example for the complete process of behaviour analysis and behaviour generation. Depending on the approaches pursued, the kind and complexity of information required for processing greatly differs. This influences the requirements on a functional mark-up or representation language.

In order to realize communicative behaviour, first of all the communicative intent underlying the behaviours needs to be computed. To do this in a principled way requires a good deal of understanding of the motivational aspects of human behaviour, i.e., why a human individual (re-)acts in a particular situation in a certain way. This requires theoretical insights into the underlying mechanisms that determine the mental, affective and communicative state of the agent. From psychology and social sciences we have a variety of evidence that human behaviour is influenced by such factors as cultural norms, the situational context the individual is in, and the personality traits and the affective system of the individual. All of which are huge areas of research where a variety of models and theories for sub-problems exist, but where we are still far from modelling the big picture of how different aspects relate and which mechanisms interoperate in which way(s). At the same time, we aim at building ECA applications with characters that display human-like (communicative) behaviour as naturally and believable as possible. In other words, we have to smartly simulate human-like communicative behaviour, which requires shortcuts at various levels of processing. E.g. somewhere in the system it is stipulated that, given certain context parameters, some character X wants to express some fact Y in a certain mood Z. Such an internal state of the system can be achieved by more or less complex processes. To which extent these processes influence the inventory and the mechanisms required for the FML still needs to be discussed.

This directly 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 their 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 requirement for reusability of components touches a crucial aspect of system and application development. Current ECA systems are built in order to realize very specific applications. Accordingly all processing components are geared towards optimally contributing to achieve the goals set out by the application. In our understanding, this is one of the major reasons why every group and almost every new ECA project has a demand for and thus creates their own, very specific representations. As a consequence the successful development of representation languages that will be shared and further developed in the community strongly depends on the ability to develop core processing components for ECA systems that are flexible enough to be customized for use in different applications and systems, and even more important that the customization process of such components provides a clear advantage over the new development of specialized ones.

Summing up, we believe successful development of representations that have a chance to be commonly used must be flexible enough to allow, on the one hand, in depth representation of theoretical insights into specific phenomena and, on the other hand, provide an inventory of high-level representations of core information that is basic to all systems generating communicative behaviour. The availability of reusable processing components that operate on this core is expected to foster the uptake of the representation language within a wider community. These considerations equally apply to the ongoing work within the SAIBA [1] initiative on the development of a common behaviour mark-up language (BML) [2] as well as

to the newly started endeavour of the development of a functional mark-up language (FML) for the generation of multimodal behaviour.

In the remainder of the paper, we will start discussing a potential inventory of an FML from the point of view of two major building blocks of communicative events, namely the communication partners and the communicative acts.

3. Some Basic Building Blocks to Realize Communicational Intent

Two basic units associated with a communicative event are the *communication partners* involved, and the *communication act* itself. See Table 1 for a tentative list of aspects of person characteristics. The listed characteristics roughly relate to three dimensions: 1. person information, such as naming, outer appearance and voice of the character; 2. social aspects, including the role a character plays in the communicative event, but also including the evaluation of a character by the others based on the outer appearance of a character, its gender, and with which voice the character speaks; 3. personality and emotion. All this influences how an individual (re-)acts in a certain (communicative) situation. Even though it is not yet sufficiently understood how these aspects interrelate to generate communicative intent, in almost all current ECA systems emotion plays an important role in intent and behaviour planning as well as in behaviour realization.

In particular, appraisal models (Ortony et al. 1988) have shown to be well suited for intent planning, basic emotion categories (Ekman 2003) are widely used when it comes to facial display, and dimensional models of emotion have been successfully employed in speech synthesis (Schröder 2004). Personality models have been integrated in agents to model behaviour tendencies as well as intent planning (e.g. André et al, 1999). The Five Factor Model of personality (McCrae & Costa, 1996) is widely used in most of the works. The interplay between personality and emotion has been studied. Ortony 2003, for instance, considers personality to ensure coherency of reactions to similar events over time.

Thus, information on the emotional state of the communication partners is important for planning and realization of the communicative acts. From an emotion theoretical point of view, a distinction between emotion proper, interpersonal stance, and general mood of an agent should be possible in the representation language, as well as a distinction between emotion felt and emotion expressed. Due to culturally dependent display rules, individuals will display different emotions depending on the current social and situational context. A clear separation between the role of emotion in intent planning versus behaviour planning, however, is not easy to draw, and depends on the power of both the intent and the behaviour planner. Some behaviour planners will be able to make use of different aspects of emotion other ones will only be able to handle emotion at utterance level.

Looking at a communicative event from a dialogue perspective (cf. Table 2), we have a structuring of the dialogue into turns, and a turn into individual communication acts. Communication acts are either verbal or non-verbal. The verbal communication acts are assigned with dialogue acts in order to specify communicative intent, e.g. ask, inform, explain, refuse, etc. As for the non-verbal communication acts communicative intent can be specified via backchannel functions such as keep contact, signal understanding, agree, disagree, etc. For an FML the question arises to which extent functional labels of verbal and non-verbal communication acts overlap and where the representational inventory differs. At the level of communication act different strands of information come together, such as information on the sender/receiver, on the emotion expressed, on the communicative intent in terms of dialogue acts and backchannel functions, as well as on information structure in terms of links to the previously communicated information versus providing new information. All

this has a potential to be encoded in FML, core aspects of which we have listed in the following tables.

Table 1: Aspects of Person Characteristics – An Initial List for Discussion

Property	Description
participants	Collection of personal descriptions of all individuals (characters) that take part in the communicative event.
person	Description of an individual taking part in the communicative event, including a unique identifier and a nickname of the character.
realname	Specifies the real name of the character. Useful in cases where real humans are represented by avatars, and the connection to the real person still needs to be kept.
gender	Specifies the gender of the character. Gender may have various implications on the behaviour of the character itself and on how the character's behaviour is interpreted by the communication partners.
type	Specifies whether the individual represented by the character is a human or a system generated character. Useful in a mixed environment where user avatars and system agents interact.
appearance	Determines the graphical realization of the character, i.e. how the character looks like, how it dresses, what the neutral posture, the base-level muscle tone and velocity of the character is.
voice	Determines which voice should be used for the character in speech synthesis and what the basic prosody parameters are, such as pitch level and speech rate.
personality	Determines the personality type of a character. The labels and values used depend on the personality model employed, e.g. extroversion, neuroticism, agreeableness in case of a simple factor model, but also labels such as politeness and friendliness may be useful in certain applications. Depending on the underlying model, values may be represented by labels or via integers or floats.
role	Role is a domain-specific attribute of the character and determines the specific role the character plays in the given application, such as buyer or seller, pupil or teacher, bully or bullied, husband or wife, mother or child, story teller or hearer etc. Thus role has a variety of (implicit and explicit) social implications which may be explicitly specified in the FML or modelled inside a processing component.
emotion	Depending on the emotion theory (such as dimensional model, appraisals, emotion categories) the representations of emotion differ. As a starting point for emotion representations related to the three different models see the work on the emotion representation language EARL [3].
emotionFelt	Kind and intensity of emotional state of the character.
emotionExpressed	Kind and intensity of emotion displayed. Felt emotion and displayed emotion are not necessarily identical, cf. display rules.
interpersonalStance	How the affective relation to the communication partner is.

mood	How the base-level affective state of the character is.

Table 2: Aspects of Communication Act – An Initial List for Discussion

Property	Description
turn	A turn comprises a sequence of communication acts of one speaker. Turns are the main building blocks which describe how the dialogue is structured.
communicationAct	Specifies a communicative act (as opposed to a non-communicative act). This may be a verbal or a nonverbal act, each of which has a communicative function or goal, and can be colored by emotion. Note, because of the embodiment of ECAs verbal acts inherently contain bodily aspects. A communication act can be a reaction to some other communication act, and it can introduce new information to the dialogue. A communicative act has its underlying producer-side intentions and goals, such as provide or get information, improve relationship, maintain or gain power, cheat, lie, etc. All these may require generalized high-level representations as well as theory-dependent in-depth representations.
dialogueAct	Refers to a verbal communication act and may consist of one or more utterances. As a staring point for the mark-up of the communicative intent, models for dialogue act mark-up such as the DAMSL [4] annotation scheme can be used, but also agent mark-up languages such as FIPA ACL [5] should be taken into account. While DAMSL (and its extension SWBD-DAMSL, [Jurafsky et al. 1997]) is a high-level framework that has been developed for the annotation of human dialogue, FIPA ACL has a defined semantics for each communicative act that is exchanged between software agents. In practice, however, for concrete ECA applications additional application-specific labels may be useful.
informationStructure	Looking from a high-level and coarse-grained perspective, information structure anchors what is being communicated onto what has previously been communicated (*theme*) and what the new contribution is (*rheme*). Information structure also influences prosody and thus may be a valuable input for speech synthesis [Baumann, 2006].
nonVerbalAct	A communication act that entirely consists of nonverbal behaviour. Typical non-verbal acts in communicative situations are backchannels. The functional labels from Elisabetta Bevacqua's feedback lexicon could be a good starting point here.
producer	Who the producer of a verbal or nonverbal act is.
addressee	Who the addressee is. Producer, addressee and hearer refer to the persons specified in the participants list of the

	communication event.
receiver	The individual who feels addressed by the producer's utterance or nonverbal act. Receiver and addressee are not necessarily identical.
perceiver	The overhearer or onlooker of a communicative act. Perceivers in contrast to receivers do not feel affected by the communicative act. Producer, addressee, receiver, perceiver are the communication act side of person characteristics.

4. Further Challenges: Separation of Intent and Behaviour Planner

Apart from coming up with a selection of properties to be specified in FML, we suppose that one of the major challenges for the specification of an FML is how much freedom the specification leaves in terms of interconnecting behaviour planning and intent planning. Consider the problem of deciding whether to use a non-verbal act such as an iconic gesture to convey a certain intention. This could, for example, be a good solution in a situation where the addressee is busy talking to someone else, where it would be impolite to interrupt due to cultural or social restrictions, and where the agent would prefer not to wait with the communicative act until the addressee has finished the other conversation.

If completely independent planning components are assumed, a rather detailed semantic description of the content to be communicated and of the situation the agent is in is required. Since FML should not contain information on the physical realisation, and if intention planning does not get feedback from behaviour planning, the component has no knowledge whether there is a certain gesture available to the agent that will serve the communicative intention. Thus the behaviour planning component needs to receive input in a detailed enough semantic description that allows for the decision that a) it would be good to use a gesture in the current situation, b) there is a gesture that conveys the meaning of the message such that no essential information is lost. In contrast, a system with less distinct boundaries between intention and behaviour planning would require less detailed semantic descriptions. For instance, given the intent planner has access to the gestures available in the system, the intent planner would be able to decide to use a certain gesture in the moment it defines the agent's intentions. Thus there would be no necessity for further serializing the information, reading it in and interpreting it inside the behaviour planner.

In practice, not every system will be able to provide or process detailed semantic information as may be required by a strict separation of intent and behaviour planning. This may be due to the real-time requirements of ECA systems, a lack of a suitable semantic representation language, or the lack of suitable and efficient semantic processing components.

The success of FML within the ECA community, thus, is also likely to depend on how much - or how little - it enforces the specification of semantic descriptions: on the one hand leaving enough flexibility to remain usable in systems that do not make use of detailed semantic representations, and on the other hand providing enough semantic detail to ensure interoperability between conforming components.

Literature

[Andre et al., 1999] Elisabeth Andre, Martin Klesen, Patrik Gebhard, Steve Allen, and Thomas Rist. Integrating models of personality and emotions into lifelike characters. In Proceedings International Workshop

on Affect in Interactions. Towards a New Generation of Interfaces, 1999.

[Baumann, 2006] Stefan Baumann. (2006). The Intonation of Givenness - Evidence from German. Linguistische Arbeiten 508, Tübingen: Niemeyer (PhD thesis, Saarland University).

[Cassell et al., 1999] Justine Cassell, Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H. and Yan, H. (1999). "Embodiment in Conversational Interfaces: Rea." Proceedings of the CHI'99 Conference, pp. 520-527. Pittsburgh, PA.

[Ekman, Friesen, 1969] Paul Ekman, P. & Wallace V. Friesen. 1969. The repertoire of nonverbal behavior: categories, origins, usage, and coding. Semiotica 1: 49– 98.

[Jurafsky et al. 1997] Daniel Jurafsky, Elizabeth Shriberg, Debra Biasca. Switchboard SWBD-DAMSL shallow- discourse-function annotation coders manual, draft 13. Technical Report 97-01, University of Colorado Institute of Cognitive Science, 1997.

[Krenn 2003] Brigitte Krenn. The NECA Project: Net Environments for Embodied Emotional Conversational Agents Project Note. In Künstliche Intelligenz Themenheft Embodied Conversational Agents, Springer-Verlag, 2003, p. 30-33.

[Matheson et al., 2003] Colin Matheson, C. Pelachaud, F. de Rosis, T. Rist, MagiCster: Believable Agents and Dialogue, Künstliche Intelligenz, special issue on “Embodied Conversational Agents”, November 2003, 4, pp. 24-29.

[McCrae, Costa, 1996] Robert R. McCrae, Paul T Costa, Jr. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), The five-factor model of personality: Theoretical perspectives (pp. 51-87). New York: Guilford.

[Nijholt, 2006] Anton Nijholt. Towards the Automatic Generation of Virtual Presenter agents. In: Proceedings InSITE 2006, Informing Science Conference, Salford, UK, June 2006, CD Proceedings, E. Cohen & E. Boyd (ds.).

[Ortony:2003] Andrew Ortony. 2003. On Making Believable Emotional Agents Believable. In R. Trappl, P. Petta, S. Payr (eds). Emotions in Humans and Artefacts. MIT Press 2003.

[Ortony et al. 1988] Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press (1988).

[Rehm & Andre, 2005] Matthias Rehm and Elisabeth André. From chatterbots to natural interaction - Face to face communication with Embodied Conversational Agents. IEICE Transactions on Information and Systems, Special Issue on Life-Like Agents and Communication, 2005.

[Schröder 2004] Schröder, M.. Speech and emotion research: an overview of research frameworks and a dimensional approach to emotional speech synthesis (Ph.D thesis). Vol. 7 of Phonus, Research Report of the Institute of Phonetics, Saarland University.

Web Links

[1] SAIBA http://www.mindmakers.org/projects/SAIBA

[2] BML http://www.mindmakers.org/projects/BML

[3] EARL http://emotion-research.net/earl/

[4] DAMSL http://www.cs.rochester.edu/research/speech/damsl/RevisedManual/RevisedManual.htm

[5] FIPA ACL http://www.fipa.org/specs/fipa00037/SC00037J.html