László Hunyadi: Multimodal Human-Computer Interaction Technologies
Argumentum 7 (2011), 240-260
Debreceni Egyetemi Kiadó

Working paper

László Hunyadi

Multimodal Human-Computer Interaction Technologies

Theoretical Modeling and Application in Speech Processing[1]

Abstract

Robots have been around for several decades with an ever increasing role, especially in industry. Nowadays they are used in information systems as well, e.g. public real-time dialogue systems. In order to assist humans in their diverse everyday needs, certain important steps are being made to create so-called smart robots. The more we wish robots to have human-like behaviour the more it becomes essential to study the nature of human-human communication in order to identify and possibly implement its major systemic characteristics in the technological world of robotics. This paper presents the outlines of a multimodal theory of communication that is aimed at both capturing the technologically relevant structure of human-human communication and offering a way of mapping such a structure onto technology. Since communication takes place simultaneously as a process of analysis and synthesis, we propose a two-way generative model assumed to be suitable to be equally adopted in both directions. Inspired by the basic approach of generative linguistics, the model wishes to meet the challenge of offering an interface between the qualitative multimodal features of communication and their quantitative representation in technology.
Keywords: human-computer interaction, language technology, human-human communication, arbitrary inter-personal communicative event, text-based interactive systems

1 Introduction – Communication from a technological perspective[2]

1.1 Human-computer interaction and multimodality

Evidently, the computer has become a crucial part of our everyday lives. A great number of users experience both its advantages and limits. Search engines, databases, and spell checkers have infiltrated our lives to such an extent that we hardly notice their benefits. We frequently draw on these Internet-based services during our daily tasks and duties such as report writing, filling out our tax returns, and sending gifts. During these processes we communicate with the computer in particular ways. Although we adapt to the computerized environment during this communication, the widespread use of these (Internet-based) services has been hampered by

[1] This research is being carried out within the TÁMOP 4.2.2-08/1/2008-0009 project: Theoretical fundamentals of human-machine communication technologies. In some respects, it is related to the OTKA NK 69042 project.

[2] I would like to thank several people, mostly members of the HuComTech team or associated with it, for offering me the chance of inspiring and valuable discussions on many of the difficult issues at the crossroads of communication and technology, especially Ágnes Abuczki, Alexa Bódog, Olga Bársony, István Csűry, Géza Husi, András Kertész, Gábor Nagy, Enikő T. Németh, Csaba Oravecz, Kinga Pápay, Tapio Seppänen and Tamás Váradi.

László Hunyadi: Multimodal Human-Computer Interaction Technologies
Argumentum 7 (2011), 240-260
Debreceni Egyetemi Kiadó

the fact that human-computer interaction significantly differs from interpersonal communication (that we have so far been used to). If this interaction becomes more human-like, the efficiency of the background services may also be significantly increased. From the perspective of the providers of these services, economic efficiency might be achieved if such systems need to rely on human intervention to a lesser extent so that the human end-users do not emphatically have to adapt to the conditions of the machine-like environment (that was introduced decades ago and has undergone considerable development since then). Users should have the impression during these interactions that they are in contact with another human, not a machine.

The first interactive systems not based on written (text) were based on speech processing and speech synthesis. These information systems restricted themselves to keyword recognition (within a certain field) and standard reply generation based on the recognized keywords (Jelinek 1976, Bakis 1976, Bahl et al. 1981, Bahl et al. 1983). Many years had passed until it became widely recognized that human-computer interaction is a much more complex phenomenon: the human behaviour of machines must mirror the multi-channel, multimodal nature of communication. In addition to the recognition of speech (processing of isolated words), it is necessary to detect and represent further, non-verbal modalities as well, such as gestures (especially the movements of the hands, head and the torso) and facial expressions (with special regard to eye and lip movements). This approach has at least three advantages. First, the joint presentation and processing of verbal and non-verbal modalities gives a much more realistic impression; second, the considerable redundancy of these modalities enables error correction, the identification of the pieces of information that have been insufficiently or ambiguously mediated through one of the channels; and last but not least, it may facilitate the emotional accommodation of the computer to the human user, a necessary condition for the maintenance of a successful human-computer interaction. Therefore, the realization of multimodal human-computer interaction is such a complex task that the simultaneous processing of information – which is mediated through multiple communication channels – requires the representation of both verbal and non-verbal communication. Theoretically, there are two fundamentally different but equally possible approaches to achieve this goal: we may either apply a descriptive, database-driven statistical, heuristic approach by which we acquire the “copy” of the observed and documented piece of reality that has earlier been organized in a goal-oriented way; or we may follow a theoretical model whose departure point is the invariant basic structure of communication that can be applied to any arbitrary context, and, based on this basic structure it arrives at the technological realization of a given event. The selection of the appropriate approach is not an easy task. The practical advantage of the first one is straightforward, since it virtually dominates today’s practice of engineering. However, it bears in itself its weakness as well. By ‘copying’ a given piece of reality and modifying it to our purposes, it always remains on the level of uniqueness without the power of generalizations owing to its descriptive and, based on the given descriptions, statistical nature. On the other hand, the difficulty of the latter approach arises from the fact that we must grasp in a generalized theory the essence of communication and in doing so we seemingly deprive communication of its main feature: the knowledge about the uniqueness and irreproducibility of a given piece of reality (i.e. event). (For the two principally different approaches see Jurafsky 2004). Since a communicative event is manifested in interactions and involves a great variety of tools of interaction, it is not obvious that we can view communication as an abstract, formal structure without any performative purposes. Most communication theories do not undertake to follow this view, rather, they describe the functions and pragmatic correlates of communication. No wonder that a variety of different paradigms of communication theory have evolved. There are few approaches that aim at unveiling the formal basic structure of communication while also focusing on the model-based (as opposed to statistical) technological implementation of an arbitrary communicative event. (For text-based – non multimodal – communication approaches see Polanyi 1988, 2004, Thione et al. 2004a, 2004b). The fact that we have set the multimodal structural description of communication as our goal, makes our task even more complex as we must reconcile the formal description of texts (a linguistic model) with a similar description of non-verbal modalities (a non-verbal model) and we must incorporate them into a single augmented model. As a result, our augmented model should be as "accurate" as a statistical model based on pattern matching but, with the power of theoretical generalizations should go well beyond the heuristics an always finite database would offer.

1.2 Technological challenges in modeling

Currently, structural modeling of communication is further complicated by the difficulties in technological implementation. During the interaction, the computer facing its human partner lacks the intelligence of its partner and possesses only a limited amount of knowledge, just the amount that we, humans, were able to formulate and represent. Although an arbitrary interpersonal communicative event between at least two people may follow certain stereotypes or constraints, intuitions and individual features of the conversation partners play a crucial role in the progress and success (or failure) of the interaction. Despite all sorts of programming tricks (such as the randomization of stylistic variations of reactions and interactions, the spontaneous nature of human-human communication) modeling can only be achieved schematically in a rather limited way in human-computer interaction. At present we confine ourselves to the exploration of the general features of interpersonal communication that must be performed during any communicative event. The non-observance or violation of these general features would result in incomplete, deficient communication with unacceptable structure, despite the fact that its smaller segments might look acceptable.

Therefore, our task is to provide technology with a general communicative structure that meets the minimum requirement: it should have significance beyond the particular pragmatic differences of arbitrary communicative events. However, while making generalizations on the basis of particular events, this general underlying structure of communication must also be extended by functional and pragmatic features in order to enable us to make a 'real life' representation of the abstract skeleton of such an event.

At this level, we may regard the given interaction minimally lifelike and human-like. In this bottom-up approach, building on the universal building blocks of communication, we move from the general towards the individual, creating a possible set of event structures by a mechanism that is able to distinguish well-formed communicative structures based on their content.[3] The basic motivation behind our model is the facilitation of such a possible turn in the relation between theory and technology. In contrast to the optimally dominant practice, according to which the technological realization of a communicative system is based on the descriptive processing of large-scale databases representing real pragmatic contexts, a model

[3] Although the concept of well-formedness is commonly used in mathematics and linguistics, we find it extendable to our multimodal communication model. Well-formedness in this case means that a communicative event together with all its modalities meets our intuitions. This recognition leads to the account of our intuitions in a formal model.

progressing from general structure to unique surface representation as well as from surface uniqueness to structural generalization can also serve as an explanatory model rather than a description, and, accordingly, it can serve as the theoretical foundation of error correction as applied to the technological mapping of a communicative event that is considered possible but whose elements are not found in the database.

Of course, this model cannot exist without relying on databases either, in other words, it cannot do without the data-level description and the processing of real and individual communicative events. The purpose of this database building is not to provide a copy of the given piece of reality, restricted to its given events, but to check, correct, validate, and improve the model. Our final goal is to create the structure of all possible (including even never before observed) communicative events, after having uncovered the general relevant features of the actual communicative events contained in the database.

In order to create such a model, we first need to answer the question whether it is possible to have a structure that provides the foundations of the events that can be experienced at the pragmatic level but cannot handle the explanation of the actual content of the given events.

To answer this question, let us carry out a thought experiment. Suppose we are looking out of the window and we can see two youngsters out of earshot. Based on the visual information at hand, we can easily conclude without any doubt that they are communicating with each other. What is more, we can even get to know certain details of the communicative event. We can figure out who initiated the conversation, at what time, how active the interaction was (one-sided or balanced), how the turns of the participants followed one another (waiting patiently for the other speaker to finish his utterance, observing the norms and constraints of turn-taking, or conversely, interrupting him) (Németh T. 2003). We can also find out whether the speaker embedded any further message into the main message of his turn. Finally, we can have strong impressions about the success and completeness of the actual communicative event.

In this thought experiment we have mostly referred to the course and structure of the communicative event. Beside intensity and (spatial) movement, analysis of the visual information refers to the particular, functional and pragmatic features of the structure as well. Based on our observations about the structure, we may have well-founded impressions about the general atmosphere of the conversation and the state of mind as well as emotions of the speakers. This thought experiment supports the hypothesis that communicative events have certain basic components that are independent of their function and the given pragmatic context. These basic components are systematically organized into an organic structure, the possibility of which can be judged intuitively on a formal basis.

Relying on these findings, it is assumed that communication has a basic underlying structure which is made up of abstract components realized in sub-events. A given communicative event can be built systematically from the synthesized organization of these abstract components. On the basis of the above thought experiment, we propose that this structure be seen as multi-level, moving from the abstract, formal level to the pragmatic realization. We assume that this model has a modular structure, based on the experiences of this observation. (About the presumptions of the psychological and neuro-physiological modularity of speech acts, see Kasher 1994).

If we look at a communicative event more closely on the level of perception and analysis, it also becomes clear whether something fits in the given communicative event, either formally or based on its content (from the general, stereotypical viewpoint, or based on our knowledge about the individual participants of the interaction). On the basis of such assumptions, we must distinguish between pragmatic competence and pragmatic performance (see Chomsky 1977). In fact, the bottom-up synthesis model of the basic structure of communication (from formal structure to pragmatic interpretation) models pragmatic competence in such a way that this competence determines the well-formedness of a given event structure on the formal, functional as well as event level. In contrast, pragmatic performance – inserting the performative levels – is about the possible environments in which a well-formed structure can be appropriate.

Whereas statistical approaches can in principle rely only on descriptions of events already encountered in their full surface realization, the introduction of the concepts of pragmatic competence and performance (as well as well-formedness and appropriateness, respectively) in a model allows us to handle events that are "only" possible, i.e. that are theoretically possible but for which observed data does not yet exist. We propose that we are able to see beyond the particular, local and irreproducible events, put them into the category of possible events, and exclude those which are considered not possible either formally, functionally, or – at the event level – in the given environment.

The basis of our judgement regarding the pragmatic well-formedness and appropriateness of an event is not a sudden, inexplicable process, but the product of the operation of our mind. Although Chomsky introduced the concept of competence in order to support our grammaticality judgements, he extended it into the field of pragmatics as well (Chomsky 1955, Chomsky 1962, Chomsky 1978). The interpretation of competence in the framework of pragmatics – beside grammar – is an essential issue from the perspective of the development of the generative theory of grammar, since this view enables the application of the generative approach also in domains where we wish to grasp and explain the general, abstract nature of the use of linguistic form, beyond the surface individualities and eventualities. The experiment by Curtiss (1982) clearly points out that the two distinct forms of competence fulfill different conditions: in cases with a deficit of pragmatic competence, grammatical competence can still operate and remain intact. (Further research also supports the assumption of the existence of pragmatic competence, its distinction from grammatical competence, and the modular nature of pragmatics as a cognitive system, see Kasher 1991, Kasher 1994.)

Summing up: the challenges for the technological realization of an event in our approach are twofold: the given realizaton is expected to be as life-like as possible both in the unique form of representation and its structure, but, at the same time, it should also bear significant general – i.e. non-unique – properties in order to derive the given representation in accordance with the representation of a set of other related events in a technologically efficient way. That is why we need a model that handles both the general and the unique properties of an event: it should be based on some general, abstract properties of communication as well as a mechanism that generates unique, surface realizations. This approach is based on the approach of generative linguistics in that, similarly to building sentences from a well-defined set of underlying elements shared by all sentences using a set of rules, we assume that both the general properties of communication and the unique features of a concrete event can be captured within such a model. As a result, this model, linking the abstract skeleton of an event to its unique surface representation could then generate all the possible technological implementations and only such.

Accordingly, the proposed model is linked to technology in the following way:

(1) The fundamental features of an event are represented as an abstract, cognitively well-formed event type,

László Hunyadi: Multimodal Human-Computer Interaction Technologies
Argumentum 7 (2011), 240-260
Debreceni Egyetemi Kiadó

(2) the structure of this abstract event type is made up of separate, but modularly consistent levels,
(3) from the basic components, the ultimate theoretical-technological model produces the event structure that is possible from the point of view of pragmatic competence,
(4) finally, technology implements the event structure in accordance with pragmatic performance in order to realize surface uniqueness.

Therefore, the building of the theoretical structure and its technological implementation are closely interlinked in our proposed theoretical-technological model. The theoretical and the technological components of the model share at least three essential features:

(1) First of all, they follow the principle of modularity along the same kinds of modules,
(2) secondly, they adhere to the duality of competence and performance,
(3) thirdly, they handle the multimodal nature of real human-to-human (and, ultimately, human-to-machine) communication.

On the basis of the above interpretation of the structure of communication, the proposed model is composed of two parts: the invariant structure and the representation of this structure. The invariant structure is made up of two further modules: the invariant formal basic structure and the invariant functional extension. The representation of these is realized at the level of pragmatic extension.

The basis of the invariant structure of communication is the formal basic structure that is able to represent the abstract structure of any communicative event. The functional and pragmatic levels are based on this basic structure and fill it with content ("meaning") and user information in order to actualize the communicative structure of the basic structure represented by abstract means. The assumption of this multi-level, modularly multi-layered structure makes it possible, on the one hand, to account for the above mentioned example (that of the youngsters having a conversation out of hearing range) about the visual observation of a human-human interaction, as far as the temporal linear structure of communication is concerned. On the other hand, we are able to view a particular communicative event as a realization of a general communicative structure valid both functionally and pragmatically. Such a structural view of communication enables technology to treat the more or less different communicative events in a uniform manner, creating uniform methods and template-like realizations. In this way scenarios general enough thematically and technologically can be applied which involve the possibility of specificity, creating suitably lifelike events that have been experienced or recorded previously as well as shortening the technological process of production. The potential significance of the model can be seen from the perspectives of analysis and synthesis: it enables the automatic recognition and categorization of certain behaviors and event types, and then, in turn, the generation of the appropriate responses. This new approach to deal with technology from a theoretically generalized way can be applied to the development of the technologies of both two-way human-machine interaction (analysis-synthesis) and one-way perception (analysis). In what follows we present the main components of such a communication model.

2 The theoretical-technological model of communication

2.1 The invariant structure of communication

2.1.1 The formal basic structure

As mentioned earlier, we assume that the basic structure of a communicative event is a formal basic structure that is characterized by two essential features: it is composed of a closed set of elements that cannot be derived from any other elements. These elements are arranged into a hierarchical structure that provides the basis for the functional and pragmatic representations, to which appropriate modular levels and an operational component are added. Figure 1 presents the basic structure of our proposed model:[4]

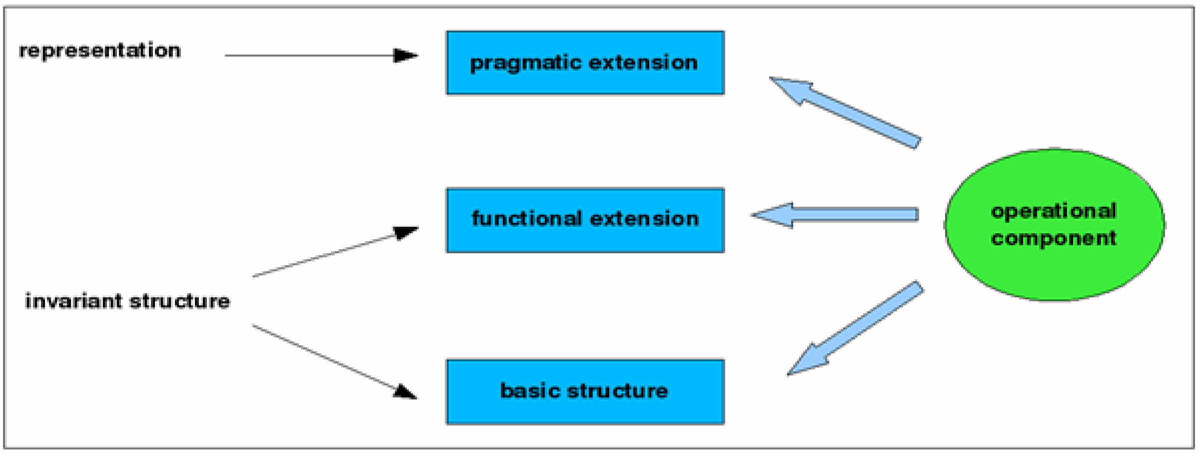

Figure 1: The theoretical-technological model of communication

The basic structure is the abstract, formal level of communication. Its basic components are the primitives which are the building blocks of all communicative events. We assume that these primitives form a closed set and the rules that organize the primitives into structure also form a closed set. These two conditions enable us to judge any communicative event independently from its function and pragmatic determinedness. As all communicative events occur in time, the primitives are temporally determined, and their arrangement follows the timeline. Therefore, it is proposed that this set of primitives minimally involves the start, end, upkeep, temporal suspension and restart of communication. The arrangement of these primitives along the timeline is the function of their relative temporal determinedness. In accordance with these, as example, we will consider a communicative event minimally possible (well-formed) if the primitive of start precedes the primitive of end, and (alternatively) the suspension and restart. We can prove this by noting that it is not possible to reverse the order; cf. (1) and (2):

(1) $CE \rightarrow b, e$
(2) $*CE \rightarrow e, b$

[4] The careful reader may notice its relation to the modular model of generative syntax as described, among others, in Chomsky 1981, but modified and extended to communication. Accordingly, the *basic structure* represents the assumed invariant formal structure ("syntax") of communication, the *functional extension* is responsible for determining the invariant functional (logical, semantic) relations and the *pragmatic extension* is where surface representation takes place. Keeping in mind the requirements of technology for sequential processes, however, the figure suggests a different, bottom-up sequential relation between the modules: the output of the basic structure serves as the input to the functional extension, and, in its turn, the output of the functional extension serves as the input to the pragmatic extension.

László Hunyadi: Multimodal Human-Computer Interaction Technologies
Argumentum 7 (2011), 240-260
Debreceni Egyetemi Kiadó

where CE = communicative event, b = beginning, e = end, and "b, e" = b precedes e on the timeline.[5]

The operational component that is applied at the level of basic structure as well as further, higher levels of the model (see 2.3) involves operations by which we structurally arrange the primitives into a structure or create the non-primitive derivations of these primitives.

The operations that arrange primitives into structure by organizing them into groups similarly operate along the timeline and involve the following: *concatenation* (serially connecting two different primitives), *iteration* (the consecutive repetition of two identical primitives), *embedding* (starting a sub-event at some point of the (initial) event, thus creating a hierarchical structure), *de-embedding* (the counterpart of embedding, returning to the higher, original – temporally preceding – hierarchical position from the embedded, lower-level hierarchical position), *insertion* (the placement of an event which is formally not a sub-event of the temporal hierarchical structure of the greater on-going event into the timeline). *Interruption* results in an incomplete structure by speaker A not allowing speaker B to complete his/her utterance. A further operation is *combination*, by which we derive groups of non-primitives from single primitives as complex building blocks of structure, without regard to their arrangement. Combination can be further applied on these derived groups along the timeline.

Although, pragmatically, a communicative event always comes to an end in one way or another, regarding its structure it is infinite since any of the above operations can be applied multiple times. This infinite process is ensured by *recursion*, the mode of the application of the above structural rules. Thus, theoretically, the hierarchical structure that can be generated from the formal, basic structure is infinitely deep. It is essential in order that we can define the set of the formally possible event structures.

Some examples of the operations that can be applied on the primitives of the basic structure:

(a) concatenation: attaching "start" and "end" along the timeline (communication starts, and it ends after a while).
(b) iteration: the repeated application of one and the same communicative primitive (e.g. the restart of an unsuccessful interaction).
(c) embedding: attaching a new "beginning" to the initial event started by the original "beginning" (speaker B, listening to speaker A's utterance, starts a new utterance instead of patiently waiting for speaker A to finish his utterance).
(d) de-embedding: (continuing the above example) speaker A returns to the suspended communicative sub-event (exactly where he finished it).
(e) insertion: it can be distinguished from embedding and de-embedding because in this case a new communicative sub-event begins and ends which cannot be interpreted along the timeline into which it has been inserted. (The result of an irregular, rather vague interaction with weak structural coherence, for example when speaker A and B constantly diverge from the topic.)
(f) interruption: the operation which interrupts an on-going event, in other words, "beginning" is not followed by "end". In contrast with embedding and insertion, the given communicative event stops without ever reaching its completion point (non-observance of joint attention).

---

[5] Although it is obvious that only one of these primitives can be present at any one time in the utterance token of the discourses (Németh T. 1996), this is also a condition for the communicative event to be formalized in the technological model of communication.

(g) combination: the combination of, e.g., primitives of hand gestures and speech prosody in variable linear arrangement to indicate turn-management.
(h) recursion, the mode of applying the above operations enables the theoretical infiniteness of a communicative event (so that we cannot determine the possible end of an event on a structural basis). This enables, among others, the deepness (and, as a consequence) intensity of such communicative events as heated arguments ("one word follows another" type of interactions).

2.1.2 The functional extension

The functional extension is the modular level where the functional actualization (in the general sense) of the communicative event is carried out in such a way that the basic structure acquires default functional values. It fills the structural frame determined by the basic structure with values that enforce the basic functional features of the communicative event. Accordingly, this module involves a finite set of functional primitives. Part of these primitives are closely linked to the formal primitives of the basic structure by determining their functional actualization. These *structural functional primitives* include: the *modes* of *start, end, suspension*, and *restart*.

A second group of functional primitives relates to the logical organization of an interaction. These *logical functional primitives* include those of *statement, negation, interrogation, condition, quantification*. They are partly realized in corresponding linguistic form, but can be manifested in more complex forms as well.

At the same time, a third group of functional primitives, the *holistic functional primitives* is not a direct functional actualization of any given formal, structural primitive, rather, they refer to the *holistic functions* of the communicative event. These primitives generally specify the interpersonal relations among the participants of the communicative event. They include: *coordination, subordination, superordination*, and such interpersonal functions as the creation and upkeep of *participation, turn-taking, continuity* (cf. Németh T. 1996), and the *emotions* related to the communicative event.

Some examples of the structural functional primitives:

(a) start: interaction between participants is initiated by one or more participants simultaneously by initiating a sequence of events that form an increasing dependency structure
(b) end: interaction between participants is terminated by completing and closing structural relations within the given dependency structure of events
(c) suspension: the further extension of the dependency structure of events is signaled to be halted by one or more participants of the interaction; the structure itself is not terminated
(d) restart: the building of a dependency structure of events is not completed but, instead of halted (suspension) the whole process is reinitiated – partly "replaying" the initial steps made before

Some examples of the logical functional primitives:

(a) statement: declaration, as a way of stating a certain position at any stage of the interaction
(b) negation: a way of continuing an interaction by stating an opposite position (opposing one's or the other's position)

László Hunyadi: Multimodal Human-Computer Interaction Technologies
Argumentum 7 (2011), 240-260
Debreceni Egyetemi Kiadó

(c) interrogation: a way of continuing an interaction by posing a question
(d) condition: a way of continuing an interaction by posing a condition
(e) quantification: a way of expressing the degree of a certain component of the interaction (e.g. an interaction can be ended smoothly or abruptly and thus its duration can be measured; the expression of the emotional attitude of speakers may also vary by some – partly individually, partly contextually – determined degree)

Some examples of the holistic functional primitives:

(a) coordination: interaction between participants in equal social or functional positions (e.g. the event of an informal conversation in which the same rules apply to all the participants of the interaction).
(b) sub-/superordination: interaction among participants who are not in identical social or functional positions (e.g. events of asking for information, recitation, disciplining, etc., where participants do not communicate according to the same rules, where one does not have the chance to mirror the other speaker's behaviour).
(c) participation: this functional primitive is crucial for successful communication. It appears in the different forms of attention.
(d) Turn-taking is a basic condition of non-monological communication.
(e) Continuity is a crucial condition of communication. It has various derived forms that can be produced by the operations of the operational component along the timeline. These involve e.g. hesitation, restart, and repairs.
(f) Emotions inevitably accompany all communicative events. We can distinguish three levels of emotions: (a) basic emotions independent of the given individual and context, (b) emotion as part of the personal profile (typical of the individual and not the context), and (c) irreproducible, unique emotion (mood) that appears in the given context, in its given form. We can regard only the first type, the basic emotions, as a functional primitive as they cannot be derived from any other primitives. The other two types of emotions are generated from the primitives of the basic emotions by the operations of the operational component. This way even the speaker's profile can be generated.
(g) Intention is also necessary for successful communication. Although some utterances indirectly aim at the formal realization of a given communicative event (e.g. interaction at an information desk) and can be formally described, other utterances can be linked to intention to a smaller extent (e.g. a question may stand for a statement, in which case it cannot initiate a question-answer exchange). Formal identification of the latter type of intention is much more complex.

Operations of the operational component which are applied on the formal primitives can be applied to the structural, logical or holistic functional primitives as well. With the help of further, primarily but not exclusively, logical operations (such as negation, quantification, and conditioning), we can obtain *derived functions*. Generally speaking: the non-primitive functions and the various levels of the functional hierarchy are derived from functional primitives by the operations of the operational component at the level of the functional extension as well. For instance, the process (and progress in time) of communication may temporarily be broken by applying the operation of insertion on the holistic functional primitive of participation realized in some form of attention. As a result of (outside) noise or some other event that does not fit in the context, attention denoting participation is also violated. In contrast to interruption, insertion ensures the re-establishment of attention, thus, the restart of participation as well. At the level of emotions, the so-called mixed emotions (such as laughing and crying at the same time) are also composed of the basic emotions, through the application of the operation of combination.

The functional realization of the formal communicative event is the frame. A frame is the functional hierarchical structure created by appropriate operations from functional primitives. Its essential feature is that, similarly to the theoretically infinite hierarchy of the possible event structure that can be formally generated from the basic structure, the functional structural hierarchy is also theoretically infinitely deep. As a consequence, while the set of functional primitives is considered finitely definable, the functional hierarchy may have a virtually infinite number of levels (hence the number of the functional modes is also infinite). It is essential that in this way we can define the infinite set of functionally possible event structures, and only those.[6] The derivatives of functional primitives and functional modes form a *set of derived functions and modes* that are considered as *possible* in an arbitrary pragmatic context. The role of the functional extension in this respect is to make these functions and modes as well as their structural properties available for the functional pragmatic selection (see 2.2.3).

2.2 Representation: The pragmatic extension. The actual representation of the functionally extended invariant structure of communication

2.2.1 The module of representation

The actual representation of the functionally extended invariant structure of communication in its final, surface output is realized at the level of the actualized, non-invariant pragmatic extension.

Pragmatic extension means the irreproducible, particular actualization of the communicative event in a particular concrete context. At this level the individual features of the participants as well as the irreproducible, particular characteristics of the communicative event must be taken into account. Although pragmatics has general, cognitive, cultural, social, moral and other aspects, the pragmatic extension is that level of communication where the singularity of the communicative event comes into prominence by emphasizing its topicality. This is the level where technology comes into effect, both in synthesis and analysis. At the same time, we may have a strong intuition that a communicative function cannot be represented by an infinite number of means. If the contrary were true, we would not be able to interpret each other's behavior and reactions.

Therefore, we assume that there exists a finite set of primitives at the level of pragmatic extension, too, and they form the basis of the final surface representation. By supposing that there are primitives at the pragmatic level of representation as well, we must also assume that pragmatics can also be described in a modular model. This model organically fits into the above outlined technological model of communication by forming an essential part of a general competence model. While this model rests on strong linguistic foundations–in the sense that the verbal, linguistic components are better documented than others, our multimodal approach enables a more comprehensive interpretation and analysis of this modularity

[6] It is indispensable to grasp the interpersonal, phatic functions of communication at the level of primitives in order to generate the event structure that is considered possible from a communicative point of view (Németh T. 1996). At the same time, recognition of these functions has a crucial role in the pragmatic actualization of the given abstract event structure.

by involving further, non-verbal modalities as well (on the modular nature of speech acts, see Kasher 1994).

Considering the modularity of pragmatic representation from the perspective of our communicative-technological model, we assume that any particular event is based on a given scenario type, the limitations of which (the possible topics, directions and turns of an interaction) are defined by the specific ontology that determines the possible nature of an event. While these features are general at the level of the type, the final surface realization of an event is carried out by setting the individual variations (with the participants' personal profile):

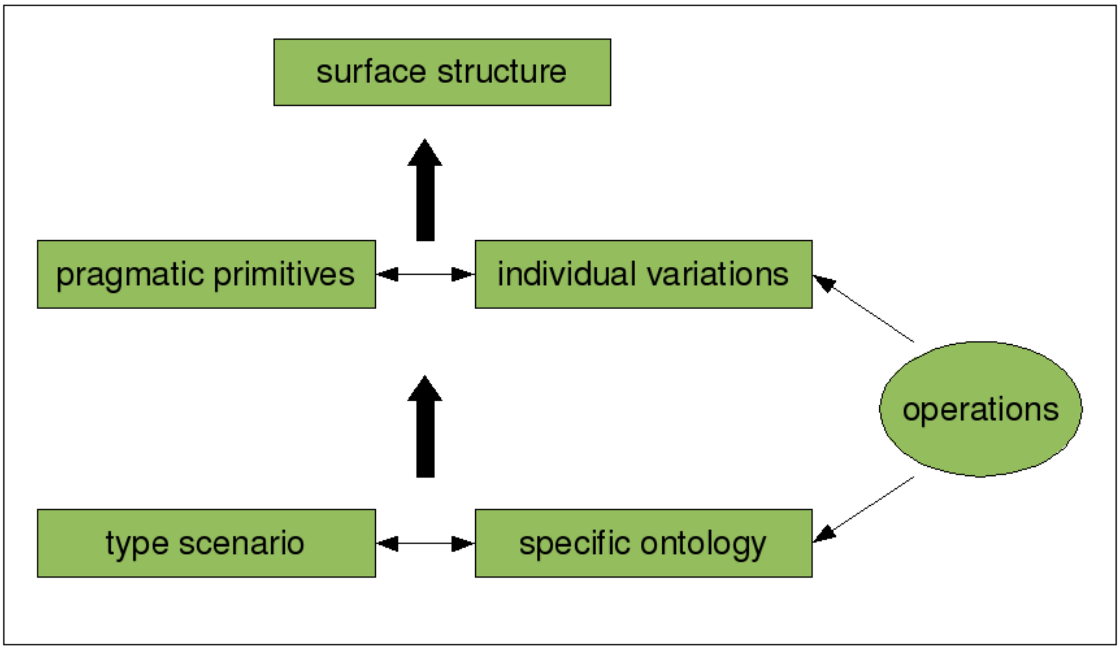

Figure 2: The representational module of the theoretical-technological model of communication: the pragmatic extension

The main components of the pragmatic sub-model of our technological communication model is described below: the two groups of the formal building blocks of scenarios, i.e. the contextually and the technologically relevant primitives, as well as the ontologies and the frame of scenarios they are applied to.

2.2.2 Primitives of pragmatic representation: the role of context towards technological implementation

An essential feature of the technological implementation that makes a context individual and unique is that it always describes a definite context. However, this particular context is not an arbitrary one, even though it is its uniqueness that first meets the eye. We know from our everyday experience that a particular event has its own general characteristics: in this way, however unique a context is, we are unambiguously able to decide whether the event matches the feature type of the given event. These feature types are described by scenarios according to which the particular context is to take place. In the meantime, a context by itself principally differs from the given scenario in that whereas the latter can be considered a predefined abstract skeleton of possible events, the context is flexible, being dynamically constructed as a function of the actual and relevant properties of the pieces of information continuously received within the given unfolding event.[7]

Similarly to other levels of the model, the basic elements of these scenario types are also primitives. Based on our multimodal approach, our assumption is that these primitives can be classified into two groups according to the two basic modalities: they can appear in non-verbal (non-linguistic) and verbal (linguistic) forms. The non-verbal primitives are *elementary movements*, *bodily* and *facial expressions* that can appear in several kinds of scenarios and contexts. Their uniqueness does not primarily stand in itself but in combination with further primitives of either the same or any different modality. We can distinguish three subgroups of verbal primitives. *At the syntactic level* it is not the meaning but the syntactic function of the word that is crucial: in this way, it is significant at the level of primitives whether the linguistic expression expresses sentence modality (declarative, interrogative, imperative, optative, exclamatory) or has certain logical relevance (assumption, quantification, negation). *At the lexical level* we find words and phrases that can be closely linked to the scenarios, and which serve the functional extension of the particular type-scenario by being the lexical means of the standard thematic realization of the type-scenario (eg. *Good morning, are you in a hurry?*, *Oh, it's hot!* etc.). Finally, *at the phonetic-phonological level*, the communicative content is expressed with the help of prosody (intonation, pitch, speech rate, pauses) and is organically associated with the syntactic and lexical forms of linguistic expression. While the significance of prosody is especially notable in verbal (linguistic, i.e. speech) expressions (cf. Jurafsky's summary on the related role of pragmatics), we can spot it also in the non-verbal realization of communication (cf. Jurafsky 2004). These primitives are related to each other by functioning under the same multimodal umbrella: whereas each of them may appear to have a particular unique function, they together, but most probably with different and variable degrees contribute to the proper representation of the given scenario.

The primitives linked to type-scenarios provide the repository of means by which, through the application of definite rules, we can generate the elementary constituents of an event (e. g. meeting someone is a complex event, but we can generate the act of greeting each other and the further constituents of the event (like the cliché-like continuation)).

Irrespective of the scenario, we distinguish two basic forms of arrangement for the event constituents: horizontal and vertical. These are generated by the model's operational component: concatenation horizontally, and embedding vertically. Similarly to the other structural levels, the operational module, through the use of recursion, ensures the creation of a theoretically infinite course of action (which cannot be defined in a finite way), but evidently only in the case of type-scenarios. The horizontal and vertical complexity of a particular realized, actualized scenario is always defined by the given action's individual characteristics. So, for instance, in the case of a job interview scenario certain questions and answers are predetermined, while in other cases it is only the questions that are predetermined, which results in the formulation of further questions according to the given answers. In this way, the depth of the concatenation along the timeline (vertical arrangement) of the given question-

---

7 Computational pragmatics distinguishes the following three types of contextual information: perceptual, discourse and mental (cf. Bunt and Black 2000), the latter including the scenario and scenario types as well. Our current proposal assumes the handling of the first two kinds of context, the perceptual context playing a role in analysis, whereas the discourse context desirable to be handled in both analysis and synthesis. Obviously, a scenario must be aware of a complex of mental information as well. This condition, we believe, however, poses a challenge for technology.

László Hunyadi: Multimodal Human-Computer Interaction Technologies
Argumentum 7 (2011), 240-260
Debreceni Egyetemi Kiadó

answer exchange, the degree of how detailed the explication is, is dynamically defined as a function of the contextual analysis of the interaction. Contrary to this, the horizontal arrangement ensures the assignment of a new topic (and also, with the help of embedding, vertically, the expression of its depth).

In this way, the complexity of an event can be described by the number of concatenated events, the vertical depth of the certain events, and the recursive generation of segment-structures (together with the embedding of the entire courses of events – such as diverging from, and returning to, the main topic).

We consider the multimodal elementary event-constituents of an event to be the constituents of a definite event, like the possible constituents of the above mentioned job-interview (and in this way it is possible to understand the given interaction as a possible job-interview) if the structural arrangement of these constituents altogether matches the job-interview-scenario. So, the basic requirement of the technological realization is that the particular action to be realized should have the hierarchical structure characteristics of a given type-scenario, and that those are in a pragmatically proper form. In order to realize these, this type-scenario has to be combined with a pragmatically possible hierarchical event-structure that is considered to happen in reality as well. In and of themselves, scenarios are unable to ensure this at present.

Scenarios – having descriptive characteristics (describing possible temporal sequence of elementary events by defining possible complex actions) – are necessary but not sufficient requirements of the type-level general understanding of an event. In order to do so, the sequence of elementary actions that is realized in the scenario has to match the knowledge that is required for the actions taking place on the level of the particular complex event. This is served by correspondence with the ontology. In simpler terms: the appropriate knowledge-segment, the shared awareness of which is necessary for the proper interaction between the participants of an event, has to be chosen out of the complete set of knowledge about the world. This knowledge, this type-ontology, will be the basis which a type-scenario can rely on, by which the unique interaction can be realized. In this way, a particular chosen scenario has to be realized by a successful technological presentation so that the scenario is also the strict result of the ontology by which it is supported. We have to emphasize that while the suggested technological communication model considers its task to define scenario types in a way that the model itself can build the basis of the realization of unique events, the construction of related ontologies itself, although assuming the ontology types, requires descriptions that fall beyond the abstract, generative frames of the present model.

The technological realization of the event structure generated on the contextual level will be the result of the theoretical model that can be perceived and experienced by the user alone. Beside the general modular structure of the model, technology also has to operate with primitives, what is more, with primitives that are organically related to the contextual primitives of the above pragmatic level of representation.

2.2.3 The technological primitives of the pragmatic representation

The role of pragmatic primitives is to create an interface between the functional representation of a communicative event and its technological implementation. The challenge is to convert qualitative data to quantitative representation (synthesis) and, vice versa, quantitative data to qualitative representation (analysis). The set of pragmatic primitives serves this purpose.

There are two kinds of pragmatic primitives: *structural* for the formal internal organization of an event, and *contextual* to actualize a given type event for a specific context.

The structural pragmatic primitives are essential building blocks of any interaction: the participants should be part of the same event by sharing (the knowledge of) the primitives of *absolute time, absolute space, referenced time*, *referenced space* and *joint attention*.

The *contextual pragmatic primitives* include *awareness* (of the given scenario and relevant ontology and used as an interface to the supporting modules of scenario and ontology) and *manner* the latter actualizing an event for the profile of the given participants.

The *principle of functional selection* represents the interface between the functional extension and the pragmatic extension. It ensures the proper selection of the derived function for the given event from among the structurally possible derived functions generated by the functional extension.

The main common property of primitives is that they are not modality specific, namely, they are general enough to be represented within any of the available and relevant modalities. These primitives are then realized by physically measurable entities, called *markers*. One and the same primitive can eventually be *realized* by more than one marker and, as one of the advantages of multimodal communication, a primitive is often realized as a combination of such markers. E.g., by characterizing an interaction as "intensive" we are able to realize this property in different modalities, and in different proportions for each modality by choosing the appropriate markers.

Markers are modality specific and in communication minimally include those related to the *visual* and the *audio signal: gestures* (represented by, among others, hand movement, eye movement, nod) and *(speech) sound* (represented by, among others, intonation, loudness, silence), respectively. Markers have sometimes apparently *fuzzy* features (cf. Zadeh 1965, Dubois & Prade 1980), such as *intensity, frequency, duration, speed, expansion, shape* which can ultimately be represented by concrete, measurable and controlled *parameters*. However, since such concepts do not have sharp, easily quantifiable boundaries, it is not straightforward for technology to handle them, even though human cognition is often represented by such “vague” concepts, with a continuous or opaque transition between two extremes, and the availability and absence of a given property. The introduction of the so called *linguistic variable* where the variable values are not numerical, instead, they are characterized by linguistic expressions (cf. Zadeh 1975) is an important contribution of *fuzzy systems* to the technological representation of similar vague phenomena and data found in everyday life. Accordingly, as an example, the concept ‘speed’ can be represented by a linguistic variable ‘speed’ where the values are also linguistic in form, such as ‘very slow’, ‘slow’, ‘average in speed’, ‘very fast’ (cf. Kóczy, T.L. & D. Tikk 2000). If we want to represent these linguistic values by technology, we have to assign to them numerical values. These numerical values are values of technological parameters, in our example those of the parameter ‘speed’. However, in order to define each of the values we have to recognize that the parametric value corresponding to a linguistic variable needs to be determined as a range or distribution of values rather than discrete ones and, since a linguistic variable such as ‘speed’ has linguistic values that are psychologically relevant, we need to determine their range by carrying out appropriate experiments. Accordingly, we will only be able to determine the parametric value corresponding to the value ‘very fast’ of the linguistic variable ‘speed’ if we discover and determine the range or distribution of the relevant parametric values based on tests with those using this linguistic expression ‘very fast’. It means that such parametric values will greatly vary depending on some relevant properties of the participants and the context of the

László Hunyadi: Multimodal Human-Computer Interaction Technologies
Argumentum 7 (2011), 240-260
Debreceni Egyetemi Kiadó

experiment. But this is exactly what we find in human communication as well: event though a given event has a fairly general formal structure, when it comes to the functions of the very same event, the actual representation of those functions strongly depends on individual and specific factors. By assuming that modality-specific markers have features that can be captured by linguistic variables and that these linguistic variables are represented in technology by particular ranges of parametric values we can arrive at our essential goal, the theoretical mapping of (aspects of) human cognition onto technology.

Accordingly, *pragmatic primitives* are the building blocks of a pragmatic structure, parameters are the minimal necessary building blocks of a technological implementation, and markers with their linguistic variable values (description as value) occupy a central and mediatory role between the parameters and the pragmatic primitives by converting and passing information between the two.

The final surface representation of communication is realized by the fact that the pragmatic markers take parameter values that are valid only for a particular moment, a particular context and participants. Although the specific values are defined by factors outside the model (such as the instructions of authors to their characters that create the given communicative event, intentions, and the individual profile of characters, for which guidelines may be provided by the scenario), the relative parameter values of the markers result from variations permitted to the extent of maintaining the coherence and interrelatedness of the different levels of the model. This makes it possible for the technology to generate pragmatically unique performative scenes while respecting the general multimodal requirements set for the given communicative event.

From the perspective of technology, the parametric values are set and modified at this level according to the pragmatic context statically or dynamically. The static (or, in other words, global) values can typically be defined for an entire given context (like stereotypes, the structural and functional definiteness of rites). The dynamic (or local) values are determined or modified as a function of other parametric values emerging within the dynamically developing event of communication.

The model suggested here is bidirectional: according to its intention and structure it supports synthesis as well as analysis. Throughout synthesis, the scenario – edited on the basis of the scenario type, on the stipulation of the technological realization at the pragmatical level – defines the selection between the possible structures that can be produced by the basic structure or the functional structure, its functional derivation, thus approaching the actualized event description of the scenario. In other words, the desired reduction of the number of possible structures to a possibly single – however complex – event structure is to be achieved by testing the formally possible structures (created at any particular level of structural generation) on the scenario and activating the most suitable one. This gradual narrowing down of the specifications of the event is followed by the setting up of the pragmatic parameters. For the final setting of the appropriate parametric values, we need to know the profile of the participants of the interaction, which fine-tunes the parametric values.

In the opposite direction the model supports analysis. Here we need to identify particular marker-values and interpret them as constituents of a definite event, so that we will be able to understand the event itself in its entire course. To do this, we have to constantly narrow down the usually expected ambiguity of the functional (parametric) interpretation of the marker values. It is the continuous choice between the possible ontology-based scenarios that plays the main role at every level of the structural analysis so that we check against such possible scenarios and activate the most suitable one as well. The choice between the possible scenarios (selecting those interpretation of the given event that are possible) is continuously narrowed down by reducing the possible structures: the more marker values we can identify and differentiate from the others – and arrange them in a single structure fitting in an ontology-based scenario, the closer we get to making the given communicative event unambiguous. Refining the unambiguity of the event is achieved by systematically taking into account the profiles and profile-values of the participants as well.

As we can see, this process consists of many components and seems to be rather complex. At the same time, in order to technologically analyze (interpret) as well as synthesize (generate) a communicative interaction, it is important to emphasize that the essentially modular nature of the model makes it transparent and technologically easy to handle. Namely, it builds on the bi-directionality and modular nature of the interaction, together with the virtually prevalent multi-level nature of communication. Doing so, it considers the scenario, the ontology and the individual profile as three interrelated and interdependent components of any event of communication. As a model of analysis, by identifying the given marker values and associating them with possible scenarios and filtering them through the interaction of the low-level system components we can arrive at the possible set of interpretations for the given event. As a model of synthesis, we follow the opposite route and narrow down the set of all possible forms of events of communication to a single, unique one by enriching the generated form with scenario-specific functions and the selection of appropriate marker values for profile-specific parameters.

3 The operational component

The operational component is closely connected to the basic structure, equally available for the levels of the functional and pragmatic extensions. It contains the operations by which, on the one hand, we can create structural groups of primitives or non-primitives from primitives, on the other hand, we can modify an entire communicative event (CE). Part of the above detailed operations (see 1.1.1) are basically of logical nature operations on sets, negation, quantification, conditional, conjunction), others are of combinatorial nature. It is essential that while there are operations that can be applied at every level of the model (like combinatorial operations), others are limited to apply at certain levels only (e.g. quantification is present only at the functional and pragmatic levels).

The task of this component is, first, to create further functions from functional primitives, second, to combine the pragmatic parameters, and assign them unique values.

Some examples of the role of operations of the operational component in the pragmatic representation:

(a) negation: with the help of negation it is possible to derive further non-primitive element-alterations from the primitives of the basic structure (for instance, to derive re-start of communication from start of communication). In the same way, on the basis of the primitives of the functional modular level it is possible to create further non-primitive functions (and modify the domain of the particular functions).

(b) quantification: similarly to negation, it is possible to apply it on the primitives of both the basic structure and the functional component level, among others, to derive degrees of functional parameters (such as the transition from one emotional phase to another) or pragmatic parameters (such as ‘intensity’ to denote the degree of a function) It may have a role in actualization at the pragmatic component level.

László Hunyadi: Multimodal Human-Computer Interaction Technologies
Argumentum 7 (2011), 240-260
Debreceni Egyetemi Kiadó

(c) conditional: with its help we can mark the possible structural, logical relations between the generated basic structural, functional and pragmatic primitives.
(d) modal operations: they can have a significant role at the functional and pragmatic levels as well. A default form of a question can be generated at the functional level, while its actualization (also reflecting profile and context) happens at the pragmatic level. This can particularly be useful in the case of some emotionally marked cases, such as exclamation and other emotive behavior.

4 The multimodality of communication and multimodality in the model

As we have already mentioned, a model has to grasp the essence of communication in order to be able to reflect it in the most appropriate way. We have to handle multimodality, a general feature as a basic structural part of the model, but the model also has to represent the multimodality of a particular communicative event at a certain level.

The duality of pragmatic competence and pragmatic performance serves this particular aim for the theoretical model. The fact that the same information in a multimodal environment is conveyed by the combination of different modalities in such a way that the relative contribution of the given modalities can vary, depending on the individual or the context (see Hunyadi 2006, 2009) suggests that the actual apprehension of multimodality is not set in the competence model (i.e. at the level of the invariant formal structure or the invariant functional processing), but in the performance-model. This is the level where it is possible to realize the surface representation of identical functional relations with a significant degree of variability.

Structurally, the basic elements of multimodality appear at the level of pragmatic representation in the instantiations of the markers. Markers can belong to several modalities and they can be modality-specific as well. Since a single function can usually be expressed by several types of markers within one and the same modality as well, the relative priority of these markers can change in each modality. Since redundancy belongs to the essence of modality, selecting the combination of markers (with a certain relative priority) to express particular pragmatic parameters is a serious opportunity for technology to represent individual differences by their variation

It is important for technology that a particular communicative event structure should appear in the form that is considered possible by the theoretical model. In this way, the most significant task of technology is to represent multimodality at the pragmatic level – mapping on it the structure created by the theoretical, invariant model.

László Hunyadi: Multimodal Human-Computer Interaction Technologies
Argumentum 7 (2011), 240-260
Debreceni Egyetemi Kiadó

5 Some issues concerning the technological implementation of the model

It can be seen that our proposed model is bi-directional: analysis and interpretation goes bottom-up from basic, non-functional structure to surface, pragmatically interpretable form, while the technological implementation of a communicative event has a top-down structure, going from physically identifiable markers and their quantifiable values to a unique communicative structure. These two directions highlight the duality of competence and performance: our model is in fact a model of pragmatic competence too, which means that in the framework of this model an infinite set of possible event structures can be generated. We claim that in an optimal case an arbitrary, hierarchical formal and functional structure generated by the model matches our intuitions about a possible communication event structure, independent of its pragmatic validity. At the same time, technology usually has a commitment to top-down processing (based on the database created from the data of real pragmatic events) and it wishes to achieve the goal that its final product corresponds to our communicative intuitions. In this case the bottom-up direction of our model fulfills the role of validation, the same way that an individual wishing to buy a product decides on doing so if the actual product meets his/her initial (general) expectations). The more data we have in our database, the better the chances that the validation will be successful.

It must be emphasized that the primitives of the model outlined do not have surface features (expressed in sounds, movements, etc.). Similarly to such linguistic symbols as *N* for 'noun' or *V* for 'verb', and their groups, *NP* and *VP*, these primitives and their groups can be seen as structural elements that can be generalized from but not identified with any element on the surface structure (NB.: however strange it may sound, we can pronounce a verb like 'see', but cannot pronounce its generalization, a *V* (verb)). The relations of primitives in a hierarchical structure enable us to judge if a given communicative event is possible without taking into account the surface realization of this basic structure or the realization of concrete communicative events. At the same time, it is indispensable to observe and describe the markers by which a given possible structure of communication is actualized and realized on the surface. The identification of these markers is indispensable in order to check and expand the set of primitives introduced in the present model. For this purpose, we are planning to carry out experiments on perception.

Identification of markers is essential from some other technological perspectives as well. Owing to the multimodal nature of communication, one and the same piece of information can be expressed through various modalities. The division of labour among them and the preferred modalities must be determined, because (1) during the technological implementation of an optional function there cannot be a contradiction between multimodal expressions of the same function – unless so designed deliberately (Hunyadi 2009, Hunyadi 2010), (2) the choice of the preferred modality marks an important pragmatic choice as well.

One more important remark must be made concerning the relationship between the model and technology. As has been emphasized earlier, our model is modular. This must be noted because we know that communication itself has a multimodal nature, namely, various modalities simultaneously interact during the implementation of a function, i.e., it might seem obvious to look at communication holistically and apply a holistic model for its description. The most frequently taken technological view relies on large-scale databases and has a statistical approach, thus, it is descriptive. The underlying reason for this is that technology does not implement functions holistically but rather step by step, progressing from one level to the next. Therefore, such a modular model can more effectively support technology whereas it is an exciting challenge to fit this modular approach into technology.

László Hunyadi: Multimodal Human-Computer Interaction Technologies
Argumentum 7 (2011), 240-260
Debreceni Egyetemi Kiadó

Although it is not feasible at this point to detail a number of problems that will probably emerge in the future, we can reasonably expect that technology can follow the basic idea of the model, namely, that data are elaborated in hierarchically dependent modules so that the output at a deeper level serves as the input of a higher level (synthesis) and vice versa (analysis). Operations with control parameters, already widely used in technology seem applicable for this purpose both in the model and the appropriate technology.

Bibliography

Bahl et al. (1981): Speech Recognition of a Natural Text Read as Isolated Word. In: IEEE *International Conference on Acoustics, Speech and Signal Processing*. Atlanta, Ga., U.S., vol. 3, Apr. 1981., 1168-1171.

Bahl et al. (1983): *A Maximum Likelihood Approach To Continuous Speech Recognition*, Mar. 1983, 179-190.

Bakis, R. (1976): Spoken Word Spotting Via Centisecond Acoustic States. *IBM Technical Disclosure Bulletin*, vol. 18, No. 10, Mar. 1976.

Bunt, H. & Black, W. (2000): The ABC of Computational Pragmatics. In: Bunt, H. & Black, W. (eds.): *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics*. Amsterdam: John Benjamins, 1-46.

Chomsky, N. (1955): *Logical Structure of Linguistic Theory.* MIT Humanities Library. Microfilm. 1955. New York & London: Plenum Press, 1975; Chicago: University of Chicago Press, 1985.

Chomsky, N. (1962): Explanatory Models in Linguistics. In: Nagel, E., Suppes, P. & Tarski, A. (eds.): *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*. Stanford: Stanford University Press, 528-550.

Chomsky, N. (1977): *Essays on form and interpretation*. North Holland, New York.

Chomsky, N. (1978): Language and Unconscious Knowledge. In: Smith J.H. (ed.): *Psychoanalysis and Language, Psychiatry and the Humanities*, vol. 3, Yale University Press, New Haven; republished in Chomsky, N. 1980.

Chomsky, N. (1980): *Rules and Representations*. Oxford: Basil Blackwell.

Chomsky, N. (1981): *Lectures on Government and Binding: The Pisa Lectures.* Holland: Foris Publications.

Curtiss, S. (1982): Developmental dissociations of language and cognition. In: Obler, L.K. & Menn, L. (eds.): *Exceptional Language and Linguistics*. New York: Academic Press, 285-312.

Dubois, D. & Prade, H. (1980): *Fuzzy Sets and Systems: Theory and Applications.* New York: Academic Press.

Hunyadi, L. (2006): Grouping, the cognitive basis of recursion in language. *Argumentum*, 2. 67-114.

László Hunyadi: Multimodal Human-Computer Interaction Technologies
Argumentum 7 (2011), 240-260
Debreceni Egyetemi Kiadó

Hunyadi, L. (2009): Experimental evidence for recursion in prosody. In: Benjamins, J., Diken, T. ten & Vago, R. (eds.): *Approaches to Hungarian*, Vol. 11. 119-141.

Hunyadi, L. (2010): Cognitive grouping and recursion in prosody. In: Hulst, H. van der (eds.): *Recursion and Human Language*. Berlin & New York: de Guyter, 343-370.

Jelinek, F. (1976): Continuous Speech Recognition by Statistical Methods. *Proceedings of the IEEE*, vol. 64, No. 4, Apr. 1976.

Jurafsky, D. (2004): Pragmatics and computational linguistics. In: Horn, L.R. & Wards, G. (eds.): *Handbook of Pragmatics*. Oxford: Blackwell.

Kasher, A. (1991): On the pragmatic modules: a lecture. *Journal of Pragmatics* 16, 381-397.

Kasher, A. (1994): Modular Speech Act Theory: Research Program and Results [collection on the foundations of speech act theory] In: Tsohatzidis, S.L.T. (ed.): *Foundations of Speech Act Theory Philosophical and Linguistic Perspectives*. London & New York: Routledge, 312-322.

Polanyi, L. (1988): A Formal Model of Discourse Structure. *Journal of Pragmatics* 12, 601-639.

Polanyi, L. (2004): A Rule-based Approach to Discourse Parsing. In: *Proceedings of SIG-DIAL '04.* Boston: MA. 17.

Németh T., E. (1996): A szóbeli diskurzusok megnyilatkozáspéldányokra tagolása. *Nyelvtudományi Értekezések* 142.

Németh T., E. (2003): A kommunikatív nyelvhasználat elvei. *Általános Nyelvészeti Tanulmányok* 20, 221-254.

Thione et al. (2004a): Hybrid Text Summarization: Combining external relevance measures with Structural Analysis. In: *Proceedings of the ACL Workshop Text Summarization Branches*. Barcelona.

Thione et al. (2004b): LiveTree: an integrated workbench for discourse processing, *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, July 25-26, 2004, Barcelona, Spain, 110-117.

Zadeh, L.A. (1965): Fuzzy sets. *Information and Control*, vol. 8, no. 3, 338-353.

Zadeh, L.A. (1975) The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, vol. 8, no. 3, 199-249.

Dr. László Hunyadi
University of Debrecen
Department of General and Applied Linguistics
Pf. 24.
H-4010 Debrecen
hunyadi@ling.arts.unideb.hu