

A Conceptual Framework for Conversations and Services

Joost Breuker

▶ **To cite this version:**

Joost Breuker. A Conceptual Framework for Conversations and Services. ELeGI Conference on Advanced Technology for Enhanced Learning, 2005, Vico Equense (Naples), Italy. pp.24. hal-00190487

HAL Id: hal-00190487
https://telearn.hal.science/hal-00190487

Submitted on 23 Nov 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Conceptual Framework for Conversations and Services

Joost Breuker
LIRMM, University of Montpellier, and
University of Amsterdam

24-02-2005

Abstract

In this article a conceptual framework is developed that specifies the terms and their relationships that will be used for specifying conversational processes in ELeGI. The first part consists of a glossary of terms, each with an informal, textual description. These descriptions are the basis for more formal definitions as to be further specified in a machine readable and re-usable ontology. This ontology, represented in OWL is still under development. This ontology, specifies the building stones for the second part: the framework itself. The framework describes at a very abstract level the dependencies between conversational processes and service providing/requesting. It is an abstract model that is to be used (and validated) in analyzing protocols of human-human conversations, and a source of reference in building conversational architectures in ELeGI.

1 Introduction

In this article I present a conceptual framework that specifies the generic structure of and phases in obtaining services. Obtaining and providing services is a form of cooperation. These services provide learning opportunities: "affordances". Besides via experiences from explicit (formal) learning services, much informal learning occurs as a side effect of the experiences and information obtained in this cooperation. Machines, by means of agent/web/grid technology, can provide services. However, these services are often too static and lack adaptivity that limit both their primary effectiveness as the learning effects. The GRID, as enabler of more complex processing and more flexible services, and the Semantic Web, as enabler of knowledge based processing, are excellent starting points to change this situation. In fact, the fusion between GRID and Semantic Web –"Semantic Grid"– tries to combine the flexible resource allocation of (stateful) service provisons of GRID architecture with semantics based services of the Semantic Web [Roure *et al.*, 2005]. However, the framework presented here is not only concerned

with the role of machine based services, but also covers the role of machines in mediating human-human services by communication tools and services. These communication tools are not only cheap and effective alternatives for telephones, but may contain facilities that enable recording and visualisation of partners in a conversation, of real or virtual environments, of documents, etc. to promote ‘enhanced presence, i.e. a setting for conversation that is enriched by new technology. They mediate a large variety of kinds of conversations that may range from goal directed, matter of fact communication - e.g. in obtaining services provided by humans –, to entertainment in efficient ‘chatting envirionments. As a side effect these tools may establish new social networks, even new communities. Social networks are important sources of information that search engines do hardly provide. In the first place, they may give access to information not available on the web: in general local or personal information. More important is the fact that trust in information and other services is primarily transferred by such relationships [Golbeck *et al.*, 2003]. In this way the participants may also obtain new information about, and insight in ‘exploiting human resources for obtaining services. For instance, the ‘enhanced presence of Buddyspace has effects on the nature of collaboration and conversation. An important role of the framework is to explain why and how a communication tool has a function and impact on these activities.

A **framework** is a generic model for a re-occurring situation: a stereotype, much in the sense of [Minsky, 1985]. The framework is a functional one, i.e. it is aimed at obtaining services in a conversational manner. I see services as a particular kind of cooperation. That means that the cooperation plays a major role in how the conversation occurs: the goal of the cooperation is in the first place the goal of a client. The framework has two perspectives: a structural one and a procedural one.

In the procedural one we will describe the usual way how obtaining services proceeds. In obtaining services, we will distinguish three stages:

1. Service discovery [1]
2. Service negotiation
3. Service performance

In each of these stages there will be a need for conversation, and the nature of the cooperation and conversation will be different in the three stages.

The structural view gives a generic description of the components of cooperation, services and conversation. The most complex of these processes is conversation. Its role is embedded in the context of the needs of participants to obtain services. Obtaining services tuned to the specific situation, needs and preferences of the client or user requires a lot of exchange of information and developping common goals. This is captured under

[1] In literature on web or grid based services the term ‘service discovery is often used to cover all three pahses. Because these stages are so distinctive in character for instance, one cannot start negotiating a service without having found a potential service provider we prefer to treat each stage more or less independently.

the notion of negotiation. Negotiation is not only the principal phase in obtaining services, but it is also the most elaborate and crucial phase of it. Therefore the structural model will in the first place reflect what happens in the service negotiation phase.

This framework is explained in Section 4. In Section 2 and Section 3 we present the major terms with which this framework is described in the form of an informal glossary. This glossary is the input for an ontology we will develop in the next phase of Task 2.6 of EleGI and will be a more formal one, specified in OWL (Web Ontology Language), the stanard for constructing ontologies for the Semantic Web. The advantage of a formal expression is not only that the consistency of the terminology can be verified, but also that it enables us to use this terminology in the specification of services and tools in the EleGI project.

The basic view is the following. There are two main categories. The first one is task centered. It deals with notions like task, agent, problem and cooperation. Here we define service as being a special form of cooperation. This is presented in Section 4.. This task perspective defines to a large extent the context or setting of the second category: the communication or conversation between (potentially) cooperating partners (Section 2). Here we define terms like turn-taking, speech-act, common ground, etc. Conversation is in the first place viewed as a necessary function in service discovery and obtaining: the ‘client should be able to negotiate services in accordance with her needs and circumstances. However, conversation may have also an important intrinsic value, as e.g. in chatting, which may have as a side effect the establishment of social networks. Therefore, the focus is in the first place on the nature of conversation. Moreover, these notions of conversation may not only be helpful in specifying and designing conversation supporting tools for human conversation, but may also play an important role in analysing, specifying and designing human computer interaction, and may also be used as a model for agent-agent communication in machines and even at the level of GRID communications. There is an additional third category which we may need to talk about things (tasks, conversations) that actually happen or have happened, such as in transcripts that may be used to illustrate use cases. This category, called occurrences, is not of central importance and is added in case we will analyze actual conversations. It is also meant to explain why we need a special vocabulary that distinguish between notions as ‘events, ‘states and ‘entities on one hand, and ‘process (‘action), ‘property and ‘object on the other hand.

2 Major concepts for Conversational processes

The focus of the conceptual framework is on conversational processes aimed at supporting two other, dominating processes: learning and cooperation. These three processes occur intertwined. For instance, in cooperation, as in all task performances, a side effect is in skill learning. This is similar to the situation of interactive learning as described by [Laurillard, 2002]. Besides practice, efficient skill learning needs coaching, advice, motivation, etc, which requires basically a conversational setting as well. A major mode of learning learning by being told is essentially based on discourse. The acquisition of ‘insight, i.e. generic knowledge occurs by discourse in the most efficient way. The term

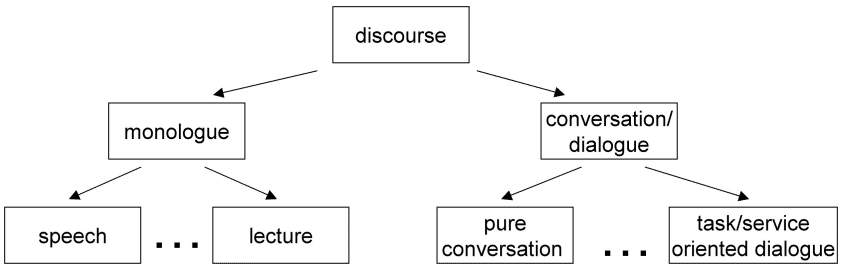


Figure 1: Types of discourse

conversational process may by itself already be a topic of discussion as its meaning is not univocal in the literature. In (psycho-) linguistics, conversation is viewed as the cradle of discourse, i.e. all forms of discourse evolve from conversation (see e.g. [Clark, 1996]). It is spontaneous and may have no other purpose then the pleasure of being engaged in an exchange of ideas and information. In general, as there is no specific goal, the issuing of topics is only locally controlled, as in contrast to functional, goal directed discourse, where the local topics can be viewed as intermediary steps towards the intended, encompassing goals. Of course, spontaneous conversation is not completely without a goal. We may start a conversation at a party with newly introduced people to satisfy curiosity, to comply with the rules of social party behaviour, etc., but this goal is insufficiently specific to guide the selection of topics for the conversation. A telephone conversation about obtaining an ADSL connection with a provider is an example of a specific goal directed conversation. In this document we will distinguish underspecified goals from intentions, which are specified, explicit, instantiated goals in plans and actions. Spontaneous, not-intention directed, free-floating conversation will be called here: pure conversation. Whether well specified or underspecified, all scholars of conversational processes (discourse) agree or assume that the use of language has never a purpose in itself. The intentions in discourse are always aimed at bringing about physical or mental changes; even if one speaks to oneself, this is not without reason. Although one may think that conversation by being the ancestor of all other forms of discourse would also be the root of a taxonomy of types of discourse, in the standard, (psycho-) linguistic jargon discourse is the most general term, covering monologues and dialogues, and written and spoken discourse. Monologues and dialogues are correlated to settings such as lecturing vs debating. Whether language is spoken or written distinguishes the type of medium used. In ELeGI we will not completely follow this convention, but replace the term dialogue with the term conversation, and use the term pure-conversation where linguistists may use simply conversation (see Figure 1)

Conversational processes can be viewed from three perspectives. The first one sees conversation as a social event, where verbal utterances are only part of a wider context of social and cultural events. This is the typical anthropological/sociological perspective. In (psycho) linguistics the focus is on the linguistic structure of discourse, while in human computer interaction, social and linguistic aspects are of secondary interest and the major concern is on the transfer of information. That will also be our (initial) focus, as the two

other perspectives may not play a major, direct role in the tool development in ELeGI. However, in analysing human-human conversation these perspectives should obviously be included.

Discourse, conversation and dialogue are concepts that are often used interchangeably. Here we will use these in the following way (see Figure 1). Discourse is the more general term and is synonymous with text, i.e. it also includes monologue and its derivatives (e.g. written discourse). If used, dialogue will be treated as a synomym of conversation (e.g. Webster dictionary). Conversation will be the most general term of interest.

Conversation is discourse in which two or more *agents* playing the *role* of *hearer*, respectively *speaker* exchange *speech-acts*.

Conversational process are the *actions* that control the exchange in *conversation*. As the exchange is aimed at some performative (cf *Speech-act*) the actions of the *agents* participating in the *conversation* can be characterized as being transactional (cf. *Transaction*).

Role Set of *norms* that specify 1) requirements on properties of an agent to take a *role* and 2) on the behaviour of an *agent* executing the role. The behavioural norms are the result of some conventional distribution of *actions* (often based on attitudes, beliefs and common goals) that imply reciprocity. The latter is the reason that roles can also be viewed as complementary (social) relationships.

Participant the undifferentiated role of an agents that actively or passively is part of a conversation.

Bystander observing role of an agent with respect to a conversation.

Speaker/Hearer complementary *roles* in *conversation*. The distribution of these roles over the participants (agents) in the conversation is controled by *turn-taking* norms (rules) of the *conversational process*. [2]

Speech-act is the basic element of a conversation. A speech act is a symbolic (verbal, gesture) *action* by the speaker intended to change the state of *mind* of the hearer. A speech act consists of two parts: a proposition, representing the content of an utterance ('message'), and a performative that indicates the *intention* with which this content is uttered (to inform (assert), to request or to order something). A more general term for performatives is *intentional attitude*.

Speaking is the action of producing speech acts. Here speaking is a generic term covering both actual speech, and gesting, but also writing. Unless specified otherwise, speaking stands for all these actions.

[2] Another term for hearer is: addressee. For written discourse the terms for these complementary roles are: writer and reader. Note that in the (pathological?) case of talking or writing to one-self, the same agent performs two roles (in talking even simultaneously)

Language is usually defined as "a system of conventional symbols for communication" (e.g. Encyclopedia Brittanica). Natural language (which includes sign-language) is here defined as a system for expressing speech acts. However, many artificial languages, including not only programming languages and other formal languages, but also communication protocols do not have means for expressing (kinds of) performatives (intentional attitudes), so we cannot define these as expressing speech acts. Note that not all artificial languages are without performatives. For instance, many agent communication languages (e.g. KQML) have means to express intentional attitudes. Any language has a vocabulary that contain symbols (terms) to refer to semantic content and a syntax that state how symbols can be combined and/or modified to form interpretable expressions.

Utterance is one or more

Process change of a property or position of an object.

Action a process *initiated by* an *agent*; the agent takes the *actor role* in an action. Actions have roles: *actor*, *objective* [3] and beneficiary (recipient).

Initiative causing an action intentionally

Actor-role role of an agent in an action; the agent takes the initiative for the action.

Intention goal of an action. All goals of actions are expected effects, and are to be taken as main, 'intended' effects. Actions (and processes in general) may have more than one effect. The intended effect(s) is are distinguished from side effects.Intentions are a 'sense-determining part of speech acts. In speech-acts two sides of intentions can be observed: those as intended by the speaker (illocution) and those as interpreted by the hearer (perlocution). In well performed discourse that establishes an maintains good common understanding, perlocution coincides with illocution.

Object role of an action is the object that is affected by an action

Pragmatics This term will **not** be used in ELeGI. It will split into two completely different concepts: dynamic semantics (see Semantics) and protocol (see Protocol). The classical idea is that a text (discourse) is built upon three layers (components): syntax ('grammar'), semantics (meaning) and pragmatics (control over discourse production and interpretation). However, it turned out that semantics and pragmatics are hard to distinguish. Modern views take pragmatics as a dynamic extension of semantics. Pragmatics are then viewed as the *discourse context* dependent modification or extension of the classical, static (truth-related) semantics. In ELeGI we will take this view as well. For specifying the rules (norms) that control *turn-taking* and *topic* management, including anaphora, which were traditionally also part of

[3]The name of this role is often: object (and actor is often 'subject'). Object and subject are rather syntactic views instead of the semi-semantic case grammar views expressed here. We reserve the term 'object' for physical or mental objects, i.e. 'real' things; not roles.

pragmatics we will use the concept of **protocol**. This is also the typical term in machine-machine communication concerning the rules for message exchange etc. In fact, pragmatics will not be a term to be used in ELeGI (for the moment).

Common ground is assumedly shared information and knowledge in discourse. 'Assumedly' means that the discourse participants hold the belief that it is shared (default) until there is evidence to the contrary. In discourse, common-ground entities can be refered to instead of being explictely predicated. Common ground consists of four resources:

Common language is the shared medium of expression of ideas (see *Language*).

Common sense is the (culturally) shared stock of knowledge. That common sense is culturally shared does not mean that it is also culturally acquired. Some basic concepts in common sense have a genetic disposition ('instinct'), like space, time, number etc., i.e. they are universal, and some are even shared with (higher) animals. (See also: semantics). Common sense is referred to by means of terms that index common sense concepts (in other words: *words*).

Common perception of entities and events of participants. The sharing is enabled by a common physical and/or social environment, plus the ability to understand the perspective of another participant of the conversation. Common perception is the basis for deixis, i.e. reference to entities that do not need description, but can be (mentally or physically) 'pointed to'.

Common discourse are discourse elements that participants have jointly experienced. Common discourse is the basis for *anaphoric* reference. The *topic structure* (see below) is part of common discourse.

Model of other participants Participants in conversation create, use and maintain models of other participants in the discourse to enable relevant, etc. discourse (cf [Grice, 1975]) one need to distinguish what is already known from what is new to a hearer. Also other information about the participants may play important roles in conversation, ranging from specific interests to social roles and status. Participants models are in HA-HA conversation what user (or personal) models are in HA-AA interaction.

Negotiating is a kind of conversation in which the participants of a conversation have the goal to establish a common ground for action. Common ground for action contains the goal, motives and required data that allow the planning of cooperation or service.

Turn-taking is a speaker/hearer role exchange. In principle the conditions (rules) for turntaking are rather simple [Sacks *et al.*, 1974]. However, violations to these rules have been observed very frequently, in particular in spontaneous conversations. As long as these violations can still be interpreted as 'being cooperative' the turn taking

is viewed as 'legal'. [4]

Topic is the referent the current discourse is about. It means that a topic has been introduced in discourse in such a way that it has just become part of the topic structure. A topic can be introduced e.g. as a comment to a previously introduced topic. For instance: "Sally met Henry, and in no time he made a fool of himself"

Topic structure is a (part-of) hierarchy of discourse topics. [5] It should be noted that the topic structure of a conversation (and in particular a 'pure' conversation) is a structure that emerges from the contributions of all participants (as long as the (common) discourse is still coherent). Topic structure is part of the common discourse of participants.

Communication Discourse is often viewed as a form of communication, i.e. as a subclass of communication. However, as the term is overused for all kinds of media, its meaning has shifted towards controlled data-transfer. We will use the term communication in this sense, i.e. if we are only concerned with the medium and the symbols with which the 'message' is transfered. For instance we can communicate a conversation orally or written; by a telecommunication medium, or by face-to-face physical presence, etc.

Protocol This term is typical for the rules that control data-transfer (communication). In ELeGI, we will use this term as well for all control aspects of discourse, i.e. the rules (norms) that control *turn-taking* and *topic* management, including anaphora.

Semantics Until recently, semantics were exclusively viewed as the meaning associated with a term (concept). However, in discourse not the full meaning of a term may be relevant. For instance, the concept 'car' has different relevant properties for the car-sales-person, than for the car-mechanic. We may say that the same concept gets different *senses*. Note that by itself the concept 'car' is not ambiguous (like e.g. 'bank'), but that it has a different meaning in different contexts. Disambiguation occurs in the same way as fixing senses. One may say that sense is the dynamic semantics of a term. However, semantics do not only cover the meaning of individual terms –even in context– but also the 'combinations' of terms in utterances.

Anaphora is making references to topics introduced in a previous part of a discourse. Typical means for anaphoric reference are pronouns and the use of definite determiners. These means oblige the hearer to find a referent in previous discourse.

[4] "The turn-taking rules only specify what is the normal or unmarked standard of appropriate behavior in turn taking." [Levelt, 1992, p 39]

[5] More or less similar to: "intentional structure".

2.1 An abstract, summarizing view of conversation: the notion of setting

The role of the Framework and its terminology is to be able to define the functions of systems existing or to be built as elements in conversation, service discovery and elicitation, cooperation, and learning. Machine mediated functions have an impact on the discourse setting. We have not defined **discourse setting**: the discourse setting involves participants, perceptual common ground, intentions of discourse, etc. In our definitions we have limited ourselves to objects and processes that are involved in discourse; not abstract, summarizing views. The notion of setting is such a view. A setting is characterized by the presence or absence of features. Introducing discourse mediating tools and other objects (e.g. pictures, maps, drawings, etc.) in the setting may change the status of these features.

Task Conversation occurs in a context. or setting. The setting is a major determinant of variations in the way conversation proceeds. The task (goal)- setting of a conversation is an important parameter. This task-setting determines the intentions in the dialogues.This task perspective is the dominant one in the framework.

Partners in conversation Besides by the task context, conversation is determined by Conversation is language based communication. In the same way as oral, natural language has been the root of diversity of discourse types, it has also been the model or metaphor by which artificial agents communicate, using specially designed languages ruled by protocols. By pairing artificial and human partners in dialogue, one can distinguish three types of communication:

HA-AA: Human agent - Artificial agent Only in the last decade human-computer interaction has deepened from machine interface design to the constructing of artificial active partners in cooperation to achieve some task [BG, 1993]. As machines have become more and more anthropomorphised by constructingagent programs that embody communicational characteristics and relative autonomy (and sometimes even some reasoning capabilities), man-machine conversation may embody elements of human communication. However, in ELeGI the aim is not to make the machine functionally more human, but to support a number of conversational protocols that should enable the user to obtain services in an intended and transparent way. Therefore, we will embed the conversational framework within this service context. A large part of the terminology includes also notions about service and on cooperation more in general. However, the ultimate goal of ELeGI is to provide an environment that facilitates learning. Learning processes are not (yet) a main topic of this document. In fact, most learning occurs as a side effect of other activities, notably conversation and (planning) task performance. These are the affordances in the learning as specified in the previous section (Section ??, based upon Laurillards Conversational framework for (e-) learning. Due to the fact

that users expect machines more and more to behave according to basic conversational postulates (see e.g. [Grice, 1975]), the design of active and knowledge-based programs are often inspired by psychological and ergonomic research and embody increasingly mechanisms derived from psycho-linguistic models of conversation: the use of user-profiles or user-models is a pertinent example.

AA-AA: Artificial agent - Artificial agent Machine (programs) can also communicate with one another, using specialized protocols and languages. The deployment of the GRID requires multitudes of dialogues between machine programs; independently whether they have agent like competences or not. In many respects these conversations can be viewed analogous to human conversation in the sense that they need an agreed upon vocabulary and a syntax. In principle, they also need some semantics for the terms used (if only in a very shallow and formal way). And they need protocols that govern the dialogue. We may see how a conversational framework is also applicable to these lower and often only rudimentary levels in which programs may talk to one another. In this document we will not make a sharp distinction between agent-like and non-agent-like machine components. In fact, the discourse competences of artificial agents are only a dim shadow of those of humans. Even in advanced AA-AA communication, as in multi agent architectures, the communications are limited due to lack of the major resource that make HA-HA conversation extremely efficient and complex: knowledge. The notion of multiple agent architecture is rather inspired by a sociological metaphor, and insect-like communication patterns than by models of human communication. It should be noted that this fact is not so much due to less sophistication on the part of the designers of these languages and protocols: on the contrary. The major reason is that the worlds these conversations refer to (i.e. the required semantics and models of situations) are very limited and well defined. They do not need to assume large amounts of common-sense knowledge to be shared, as in human conversation.

HA-HA: Human agent - Human agent Human conversation has been to a large extent the requirement in designing HA-AA communication and in some respects the analogon to model AA-AA communication. Therefore, the conversation framework as specified in Section 2.6 is primarily inspired by insights in human to human conversation. Moreover, HA-HA plays an important role in ELeGI. Humans in ELEGI are considered to be resources for offering services. In this sense, WP-6.2 on Conversational Processes and Enhanced Presence has the specific goal to end up with a specification of the role and behaviour of human to human conversations through the GRID, such as those supported by tools such as Buddyspace or Hexagone.

Other factors A well-known description of discourse setting and its features is given by [Clark, 1996, p 9.], and I will quote this because it is highly relevant in specifying the influence of tools and the availability of information in discourse.

After explaining that spontaneous conversation is the basis for all discourse types ("basic setting"), Clark argues that discourse types are dependent on discourse settings

and that these discourse settings can be characterized as presence/non-presence of the features of the basic setting.

"NONBASIC SETTINGS

What then makes other settings not basic? Let us start with the features of face-to-face conversation listed here (Clark and Brennan, 1991):

1	Copresence	The participants share the same physical environment
2	Visibility	The participants can see each other
3	Audibility	The participants can hear each other
4	Instantaniety	The participants perceive each other's action at no perceptible delay
5	Evanescence	The medium is evanescent –ite fades quickly
6	Recordlessness	The participants' actions leave no record or artifact
7	Simultaneity	The participants can produce and receive at once and simultaneously
8	Extemporaneity	The participants formulate and execute their actions extemporaneously, in real time
9	Self-determination	The participants determine for themselves what actions to take when
10	Self-expression	The participants take actions as themselves

If face-to-face settings are basic, people should have to apply special skills or procedures whenever any of these features are missing. The more features are missing, the more specialized the skills and procedures. That is borne out informally

Features 1 through 4 reflect the *immediacy* of face-to-face conversation. In that setting, the participants can see and hear each other and their surroundings without interference. Telephones take away copresence and visibility, limiting and altering language use in certain ways. Conversation over video hook-ups lack copresence, making them different too. In lectures and other non-personal setting, speakers have restricted access to their addressees, and vice versa, changing how both parties proceed. In written settings , which lack all four features, language use works still differently.

Features 5 through 7 reflect the *medium*. Speech, gestures and eye gaze are evenescent, but writing isn't, and that has far-reaching effects on the course of language use. Speech isn't ordinarily recorded, but when it is, as on a telehone answering machine, the participants proceed very differently.

. . . .

> Features 8 through 10 have to do with *control* -who controls what gets done and how. In face-to-face conversation, the participants are in full control. They speak for themselves, jointly determine who says what when, and formulate their uttereances as they go. In other settings, the participants are restricted in what they can see when. The church for example determines the wording of many prayers and responses. In fictional settings, speakers and writers only *make as if* they are taking certain actions – Gielgud is only play-acting his role as Hamlet – and that alters what they do and how they are understood. And in mediated settings, there are really two communicatiosn. Wim says "Heeft u honger?" in Dutch, which David translates for Susan as "Are you hungry?" Susan is expected to hear David's utterance knowing it is really Wim who is asking the question. The less control participants have over the formulation, timing and meaning of their actions, the more specialized techniques they require." (end of quote).

The first feature, co-presence, is not the same as what we have coined *"enhanced presence"*, but is related. Enhanced presence is not the physical presence of participants, but the awareness of their accessibility. The notion refers (a.o.) to potential participants, who are 'available. They do not participate and are neither bystanders. In fact, they are not part of a conversation at all, but may get involved when their cooperation is required. It is strongly suggested that the basic conversational setting, i.e. one that enables pure conversation, has the largest number of features present. The more features present, the more unconstrained and natural the conversation. It also suggests that machine mediated conversation of partners over physical distance from telephone to Buddyspace may only approach the ideal of physical presence. That is not the case. It is rather that technology may enable new forms of conversation. This is most obvious in the 'mediating features. Conversation can be recorded now easily (but this appears to have little effect on conversation behaviour.), but a little delay in 'simultaneity in very long distance calls has drastic effects, which means for instance that mediating software should not create noticeable delays. A reason the more to rely on sufficient computer power under all circumstances, and the GRID may be more than welcome in this area. As far as the control features are concerned, the mediation by machines should remain so transparent that there are no doubts in HA-HA conversation about the authenticity and autonomy of the partners involved: a problem that is obvious in much 'chatting software and practice.

As distance may impair the quality of a conversation setting, machine mediated conversation may add features to the 'basic setting that may improve communication and ease of understanding. First of all, participants are no longer bound to a physical setting. The setting may be a virtual one, and far more important it may enable to include objects in the 'common perception which in physical settings are difficult or impossible to realise, such as images and sounds of actual and non-actual events and objects.

This short exploration is meant to illustrate the use of a conversation framework to define, explain and predict the effects of conversation mediating and supporting software. It also gives insights how we may achieve communication between humans and machines that is in accordance with what humans expect and understand. Indeed, much work in

human-computer interaction is nowadays inspired by models of human communication. However, this is not the only reason to develop such a framework. As machines have to communicate in the Web, but in particular in the GRID, HA- HA conversation has proven to be an inspiring model for understanding how we can boost the efficiency of this required communication. As we will explain further, human communication is at a particular trade-off between accuracy and reliability on the one hand and efficiency at the other hand.

3 What has to be achieved: the notions of task and service

Discourse (conversation) is always directed towards achieving something. The elementary unit by which something is achieved is called action; the intended effect (goal) of an action is called intention. We take action here to mean the primitive "atom" from which all kinds of complex changes may be composed (as in tasks, methods and plans). All actions have a physical effect, i.e. they are achieved by human agents from some intention to some muscle(s) movement(s).

There are two major types of actions: 1) actions intended to bring about physical changes to objects and 2) speech-acts, i.e. actions to bring about mental changes. The latter are in fact constituents of discourse itself. Speech acts also need a physical medium (sound; writing) and also consists at the physical level of muscle movements. However, the analogy between speech-acts and other actions breaks (somewhat) down at a certain point. Physical actions can be virtual. A plan consists of some structure of actions, but a plan still exists, independent whether it is executed or not. I have many plans of what to do to improve my house. Some exists already for twenty years and are not executed; some others have been executed, etc. Speech-acts in discourse are also planned, but the plan itself is of little relevance: one may say that the speech-plans are rather stereotypical frames, reflecting grammar, conventional style, etc., where only the topical structure (what to say) is explicitly planned, governed by intentions (illocution). Discourse topics in speech-acts are analogous to physical objects in physical actions. Where the analogy between speech-acts and physical-actions breaks down is in their effects. Our actions may not have the intended effect. In executing physical actions this is (mainly) due to incomplete knowledge of the state of the physical world. (In most cases) we experience the effect of physical actions. However, the effects intended by speech-acts (e.g. promises; assertions) are often not directly observable, and moreover they need intermediary actions by the hearer: they need interpretation. The physical world does not need an interpreter to re-act.

The interpretation side of discourse is not simply the mirror side of discourse generation. We need all kinds of checks to make sure that we are understood and this makes up an essential part of discourse. For this reason Searle proposed to distinguish between illocution (the speakers intention) and perlocution (the hearers interpretation). When Philippe states, "The window is open", I may interpret this to be an explanation why all papers fly

around in the room, while Philippe intended me to interpret this utterance as a request to close the window. This ambiguity does not exist in physical action: nature does not make ambiguous interpretations; it is for us to interpret by observation. This all is to say that in this section we will be only concerned with those actions that have physical effects and to keep speech acts as part of an ontology of discourse. Of course, and that is why we have this introduction, at the more abstract level, discourse and cooperation will have common roots. Conversation is a kind of cooperation. The difference with cooperation consisting of physical actions is that it may be designed freely. We may plan physical cooperation in all kinds of roles (even institutional ones) and built societies out of this. However flexible the use of discourse may appear to be, the fundamental roles and actions are more or less fixed, highly conventional and deeply ingrained in language instincts [Pinker, 1994]. This limited flexibility is expressed as the notion of protocol.

In this section we will define a number of terms, such as task, goal, method, strategy, plan etc. which are often used indiscriminately: either by negligence or by the fact that in different communities different terminologies have been established. The terminology that we will propose here comes mainly from human-machine cooperation and, more in general, knowledge engineering. For an overview and more detailed descriptions and a framework, the reader is referred to [Breuker, 1994]. Most of the terms defined here concern mental objects. A plan or a task or a goal, etc. are entities that exist in the mind of the creator of these entities. Plans may contain physical actions and processes, but only if they are executed, they exist in the physical world and concern physical objects. To be precise: if a plan gets executed, the actions are instantiated into changes in the physical world.

Intentions, topics etc. are mental objects: when a discourse plan is executed by the speaker the hearer observes auditory or visual data (utterances) that need interpretation. However, different from physical actions, the speech-acts that have to be re-covered in interpreting these data involve the modelling of the mind of the speaker, i.e. need the instantiation of mental objects and processes.

Thus far we have distinguished physical actions and speech-acts. However, for modelling reasoning and problem solving we need mental actions or processes as well. The best way to explain this is to distinguish planning from a plan. In planning we may mentally assemble (representations of) actions and check whether they will achieve the intended effect in some mentally simulated physical world. This is a (mental) test to predict whether the plan (assembled actions) will indeed achieve the goal. The plan solves the planning problem. This plan is a mental object, containing physical actions. If executed, we have a real validation of our plan in the physical world.

However, we can also make plans that do not contain physical actions, but mental actions. We may plan to pay attention, etc. In fact, there is not much mental planning in this sense: the mental world is not fully under our control. Attention is easily diverted, etc. and many mental events are rather viewed as caused by mental processes (e.g. forgetting) than by mental actions. However, in conscious reasoning we have (the illusion of) control over the sequence of mental operations (actions) and this control is called a problem solving method. A problem solving method is a plan consisting of mental actions to control

the making of inferences, which are in fact the real actions that get things done in the mind, as far as conscious, explicit cognition is concerned. Mental-actions and physical-actions-as-represented-in-the-mind are hard to distinguish because in making inferences, our mind simulates to a large extent the physical (and social) world [Clancey, 1985].

Plan is a sequence of actions and processes, with an implicit or explicit structure of (sub-)plans to achieve some specific goal. The structure of a plan is a part-of structure. Plans are initiated by actions, but may also involve (cause) also processes that have predictable effects. The plan to take the train to Amsterdam includes the movement of the train itself, which is a (complex) physical process. A (structured) plan is therefore a plan consisting of sub-plans, bottoming out in (assumedely) executable actions and predicted porcesses.

Goal is a desired state of some situation. A goal is equivalent to an intention. We will use intention for the desired (intended) effect of a speech-act, and goal for the desired effect (result) of physical-actions. As actions are the building stones of tasks and plans, the aggregation of these action goals are also the goal of a plan or a task. Or rather: plans and tasks (and methods) specify actions by decomposing the task- or plan goal. The goal of a task or plan is to achieve a state that is viewed as a problem in a current state.

Problem a difference between a current situation (state) and an intended or desired state. In this sense, a problem is an occurrence (see Section 2.4.2). However, in problem solving (and question answering) methods can only be applied if there is some notion of generality: recognizing a type of problem [Breuker, 1994]. Problem types are generic concepts.

Task is a sequence of actions with a goal. That is similar to the definition of plan, and indeed tasks can be planned. Like plans, tasks are decompositions, bottoming out in actions. Tasks differ from plans because the decomposition is based upon a method, i.e. the method that justifies and explains the decomposition of a task in subtasks, bottoming out in actions.

Method is a fixed, generic plan of actions. A method is an abstract plan that can be re-used in similar situations and for similar problems. This abstraction may also involve the justification why the method should work or should yield correct, efficient or valid results (solutions).

Problem Solving Method is a method of reasoning aimed at solving a problem. The actions are therefore mental actions, in particular: making inferences. A problem solving method is a generic plan based on the decomposition of a task in sub-tasks. This decomposition entails dependencies between the sub-tasks, i.e. the output (result) of one sub-task is the required input for another sub-task. A problem solving method (PSM) needs therefore also control over the successful execution of the sub-tasks.

Cooperation is a task distribution, where different sub-tasks are assigned to different agents, acting as actors. Note that this definition states what cooperation is; not how it is achieved or performed. The *negotiating* to establish the plan for cooperation is an essential preliminary condition for cooperation. For the execution of cooperation coordination of actions and transfer of results of sub-tasks is required

Transaction is a simple, basic unit of cooperation in which two partners (agents) perform coordinated, complementary or compensating actions in which the partners play both the role of actor and of beneficiary in the respective complementary actions. A typical transaction is buying/selling of goods. The seller (actor) 'gives a good to the buyer (recipient) and the buyer (actor) compensates the seller (recipient) by paying money. The cooperation in conversation has also a transactional character in turn taking.

Service is an action in which the beneficiary has the authority to request and specify this action. The role of the beneficiary is usually called client (or 'service requester) and the agent who performs the action (actor) is called service-provider. The authority may be based on a social, even a legal position, but it may also be established ad-hoc by compensation (paying, social esteem, etc): therefore many services have the character of transaction. In modern society, buying services has probably outgrown buying products. Note that a service is valued for its result: a product such as an answer, a repair, an advice, a qualification, etc. This product may be unique and personalized. Different from simple, classical buying-selling transactions, where products are exchanged, in services we do not expect 'off-the-shelve products, but tailored actions that fit personal requests or needs. Also important is the fact that the service provider should be qualified and expert in performing the action, even if it is an advice about 'off-the- shelve-products'. Finally: a service may involve more than a simple action or transaction: it may involve long-term cooperation, as e.g. in education.

Delegation is the transfer of playing a role to another agent than the originally assigned one.

4 A Framework for Service discovery, cooperation and conversation

The framework is constructed from two points of view. The first one is a structural one, in which components and their dependencies are specified; the second one is a sequential one, describing phases in service discovery. The notation used is rather simple and informal. In the next phase of the workpackage these models will be further formalised in OWL, respectively UML formats

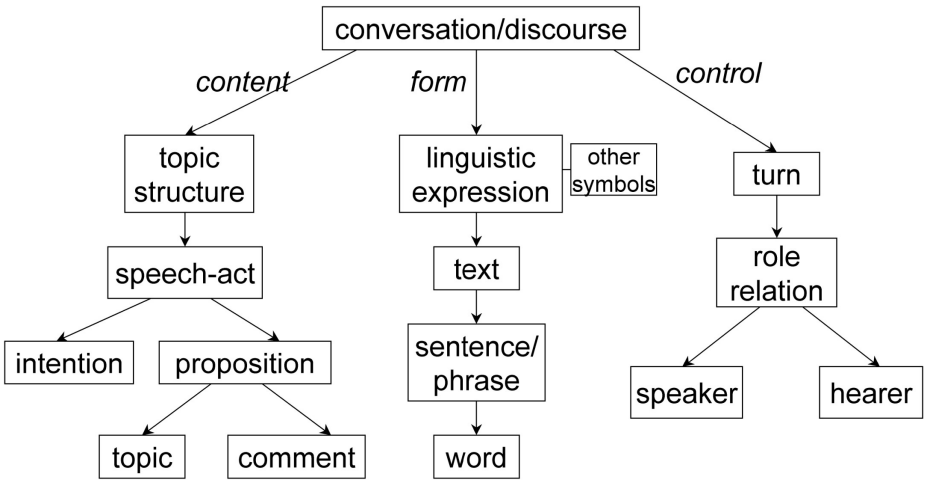


Figure 2: Components of conversation

4.1 Structural view

The structural view is specified by a series of partial models. The first models describe the composition of various types of objects and processes in conversation. Thse models are integrated in models that describe the dependencies between these objects and processes. The dependencies are expressed as input/output roles.

Objects Conversation is a complex object, consisting of a number of objects. Moreover, conversation has three views: a content, a form and a control perspective. In the ontology it means that conversation is represented by multiple classes. These multiple views are indicated in Figure 2, Figure 3 and Figure 4 by the labels of the ‘part-of arrows (relations).

In Figure 2 there are three views that more or less coincide with the classical distinction in linguistics between semantics (content), syntax (form) and pragmatics (protocol). This mapping is not perfect, as topic-structure and the topic-comment distinction is also determined by ‘pragmatic factors. As explained above, much of classical pragmatics are rather viewed as dynamic aspects of semantics which provide specific senses to the terms in linguistic expressions; those aspects of pragmatics that provide control over the discourse have been categorized under the notion of protocol.

Conversation conveys intentions that give a sense to the propositions that are part of the speech-acts. These propositions are the carriers of specific information and knowledge. Often these intentions are not explicitely expressed, but they can be derived from the setting, or more precisely: from the common-ground. More in general, the common-ground is the basis for all kinds of inferences and assumptions that are not expressed. The principle that makes natural language discourse so incredibly efficient is due to the fact that what is common and shared is no news and should therefore left out of the conversation. Therefore, the common-ground acts as an important resource to fill in assumed or implied knowledge and information. Not all information and knowledge that is used to embed the information in the discourse is common: the model that a participant has about

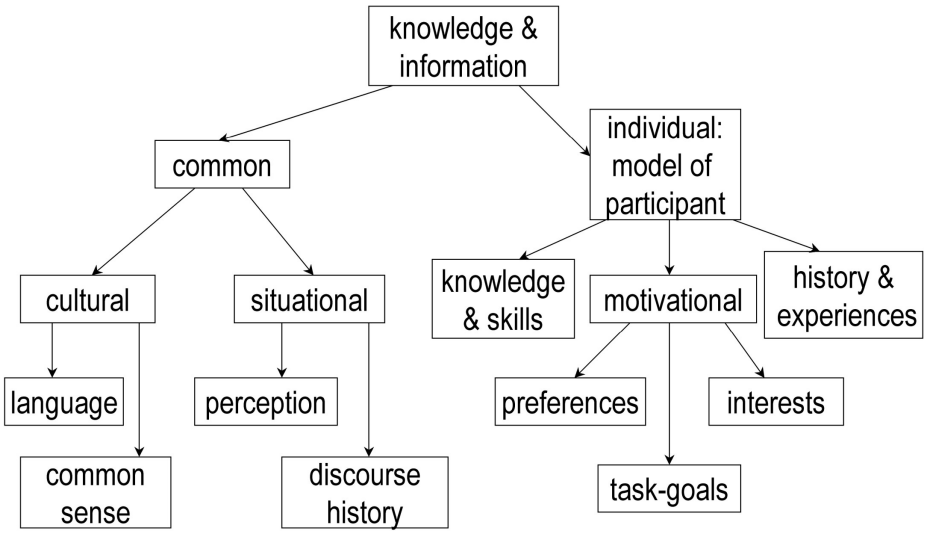


Figure 3: Components of knowledge and information in discourse (arrows indicate part-of relations)

an other participant is not necessarily shared. To some extent a partner may know (or believe to know) what the other partner knows or believes about her, but this mutuality may leave much room for individual beliefs. In Figure 3 information and knowledge objects are decomposed that can be used as resources in interpreting and constructing discourse.

Processes, actions and tasks Processes, actions and tasks represent dynamic aspects of cooperation and conversation. Strictly speaking, we should talk rather of actions than of processes, as we have reserved the latter term for changes of objects due to physical causes. In actions, the cause of a change is (in the intention of) the actor. Below we will explain that the difference can be expressed by a difference in role taking. If the actor role is performed by a physical force or other form of energy, the change of an object is attributed to a process; if this role is performed by an agent, who intends the effect and performs the required physical movements, the change is attributed to an action. However, in this document we will not be very strict in applying this distinction, as common-sense terminology may easily view aggregated actions as processes. For instance, business processes are fuelled by (mostly) human actors, but their aggregation makes them so anonimous that they look like processes that occur naturally (i.e. by their own nature).

Like objects, processes can be decomposed. With decomposition we do not mean that we descend into lower grain size levels that need new qualitative descriptions (ontology); we do not make quantum leaps. This sounds vague, but we do not decompose actions into, for instance, the neural components involved. The ontology does not include meta-levels or aggregation levels that need reification. Tasks can be decomposed into sub-tasks, bottoming out in actions. Actions are assumed to be the basic units for the planning of methods, although the ‘size of this unit may vary from domain to domain, and is dependent on levels of expertise (skill) within domains. Writing a letter of appeal is an action

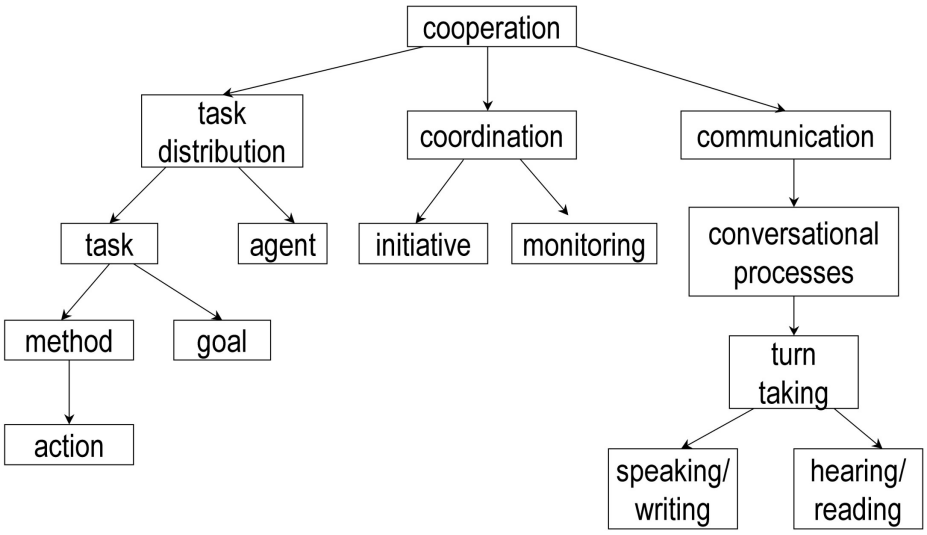


Figure 4: Decomposition of processes involved in cooperation (arrows indicate part-of relations)

for an experienced lawer; it is a task for me.

In Figure 4 the major processes in cooperation are presented. Although in EleGI service and service discovery are central themas, these terms do not appear there. The reason is that we see service as a special kind of cooperation, where the initiative and the receiving role is with one participant who is a client. Initiative and role are parameters in cooperation,

Roles, objects and actions: a framework Task and object views can be combined by specifying the roles that objects play in actions. Actions provide a frame, which consists of a core and a number of resource roles. The core of an action is the change of an object. The object has the object role, and in this role some value of a property or attribute of an object changes. When it actually happens, we may say that there is an event in which an object is in some ‘initial state and is at some later time in a ‘final state. This is the terminology for occurrences. An example may illustrate the difference between events and states on one hand, and processes and object properties on the other hand. A thing may change colour, e.g. changing from green to red. This event may be explained by totally different processes of action. The thing may be a fruit and then the change is probably explained by the process of riping. However, even if it is a fruit the change of colour me be due to some painting action. Processes and actions provide explanations for events (and also for states); they enable the disambiguation of interpretations of situations. Common-ground, in particular common-sense, is the major resource of information and knowledge for this disambiguation.

Roles in processes and actions can be simplified to input-/output roles. The input of a process consists of an object with some property and some resources, e.g. energy or information. The output consists of some effect: the change of (the value of) a property,

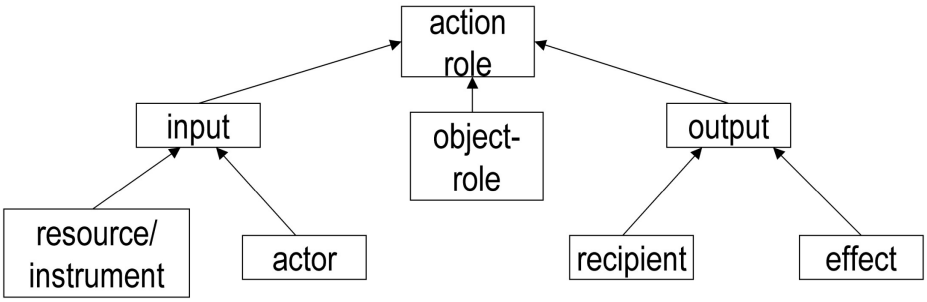


Figure 5: Typology of roles (arrows indicate is-a relations)

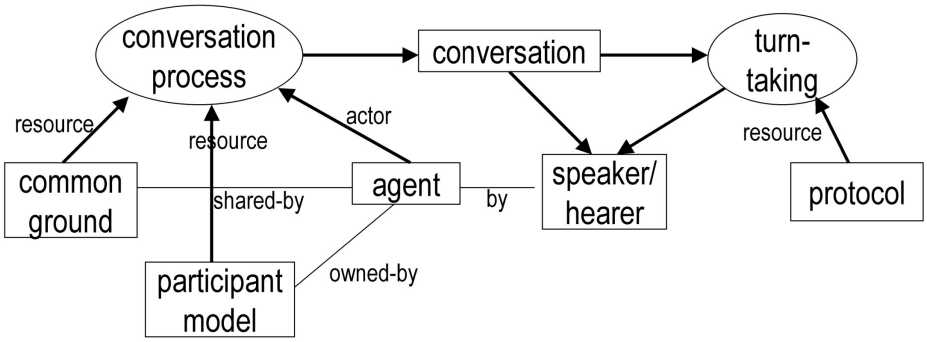


Figure 6: Dependencies in conversation processes (arrows indicate dependencies (roles)

and some consumption of resources as a side effect. As an important class of processes is concerned with movement, i.e. changing the position of an object (cf. dynamics), in information transfer the ‘new position of the object is at the hearer. We may say that the hearer has the recipient role in the action. In Figure 5 we give a simple classification of roles as sub-types of input and output roles. It should be noted that there is no clear agreement in the literature on roles in actions. On the one hand, in linguistics a a large variety of role types have been proposed, particularly as a semantic basis for the interpretation of verbs (that denote actions or processes). In many natural languages, linguistic roles are syntactically expressed as ‘cases, providing a bridge between syntax and semantics. On the other hand, software specification languages only distinguish input- and output roles, because they are used to specify dependencies between computational procedures. Here we will use a kind of intermediate vocabulary, which is classified as input-/output roles.

By linking processes by their input-/output relations, the dependencies between processes, and between processes and objects can be specified in a dependency network. Given the decomposition hierarchies in Figure 2, Figure 3, Figure 4 and the specialization hierarchy of Figure 5, we can now construct the dependency diagrams that connect processes with one another via roles filled by objects. Figure 6 represents conversation processes

The role of conversational processes is to support the planning of cooperation between two or more agents: in particular they aim at the negotiating of goals and other requirements in convening some distribution of tasks between the agents (participants).

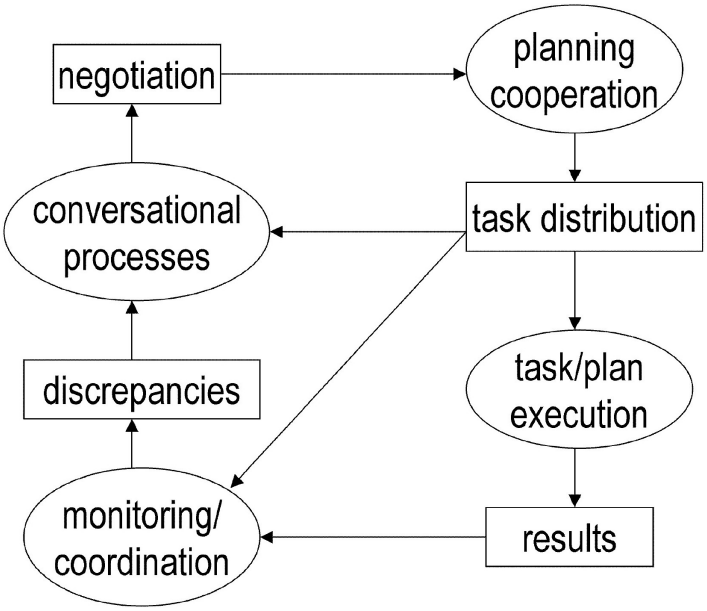


Figure 7: Main thread in cooperation, supported by conversation (arrows indicate dependencies (roles)

In Figure 7 we present the main thread and processes in planning and executing cooperation. This is a simplified, product oriented view in which we show only the inputs and outputs that contain results produced in one task or process and that are inputs for a dependent task or process. These input-output roles are called 'dynamic roles, because they are intermediary products, in contrast with 'static roles that contain resources. [6] In Figure 8 we have added two important resources: agents and common-ground. Note that common-ground not only plays a decisive role in conversation, but also that some common ground is needed to establish coordination. Coordination in the execution of tasks by different agents may proceed via conversation, but in case of a shared environment a perceptual common ground may be sufficient.

Services are viewed as a special kind of cooperation in which the roles of the cooperating agents are no longer mutually exchangeable and equivalent. In the first place, the initiative to request for a service – i.e. to engage in a cooperation is with one of the agents, which also gets the role of client. The other agent becomes the service provider. Once the cooperation is planned, the execution of the task distribution or, at least the part of the task distribution that is selected as a service is performed by the provider. In fact, in the execution there is no longer explicit cooperation in the sense that both agents execute their tasks. There is some rudimentary cooperation because the service results may be part of a larger task of the first agent, and because this agent also monitors the (results) of the task execution of the service provider. This inequality in roles has usually a price: services are traded for other services or goods: notably money. This part of servicing that indicates

[6]The distinction between dynamic and static input roles of tasks or processes is taken from CML, the conceptual modeling language of CommonKADS (see [Breuker, 1994]

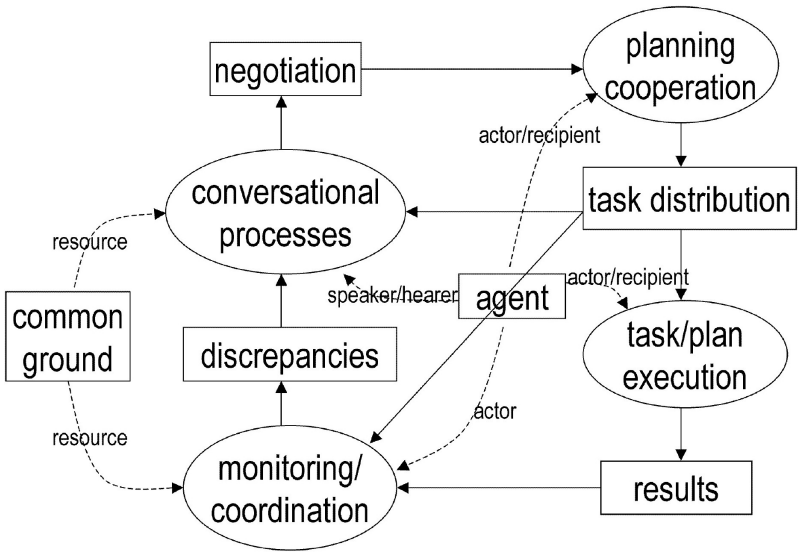


Figure 8: Cooperation supported by conversation (static roles added)

that obtaining services is rather a transaction (like exchange of goods, commerce) is not explicit in Figure 9, which is a refined version of Figure 8.

4.2 A procedural view in obtaining services

In Section 1 we have distinguished three phases in obtaining services:

1. Service discovery
2. Service negotiation
3. Service performance

Each of these phases may be simplified and reduced to simple actions. However, in obtaining services in real life, each phase may consist of complex plans of tasks and conversational actions (speech acts). Moreover, phases may not simply follow one another, but may be part of complex refinement and 'trial-and-error cycles. A major reason that obtaining services is complicated is due to the fact that the client may not have articulate needs for services. Another major reason, which is the other side of the same coin, is that the offering or discovery of available services may complicate and confuse a potential client. In this respect, the trading of services is not different from the trading of goods.

However, there is a major difference in obtaining goods, i.e. ready made products, and in obtaining services. The trading of goods can be viewed as a fixed, extreme case of trading service. As needs may be shared, it is not surprising that meeting these needs becomes predictable, and instead of complex negotiation and ad-hoc cooperation, it becomes more economical to organize the production of such good, with the side effect of quality assurance, lower costs, etc. The other side of the coin is that individualization and

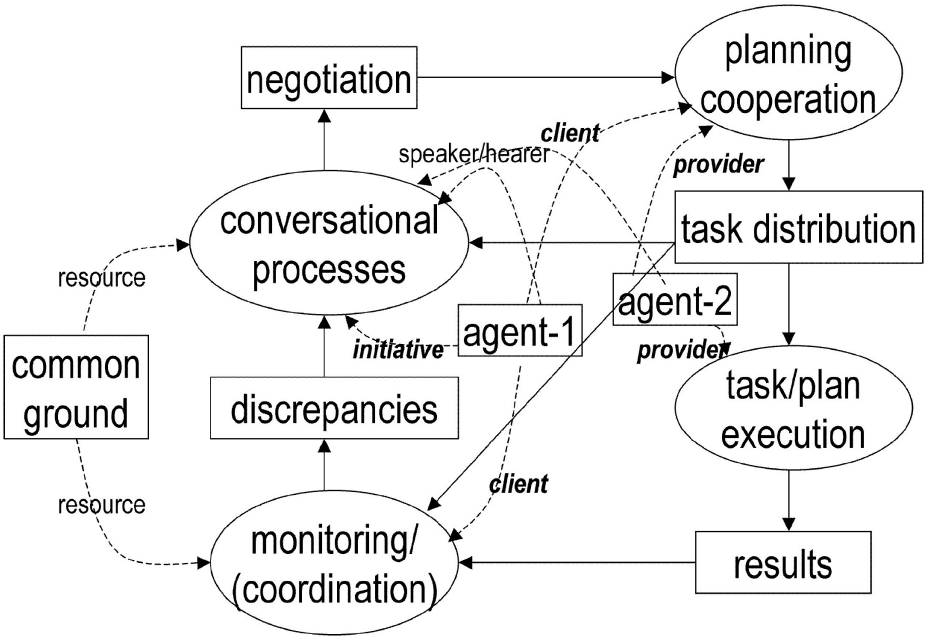


Figure 9: Obtaining services modelled as a special case of cooperation

mutual understanding of client and provider about the clients needs are lost. Service becomes a luxurious way of obtaining goods. This is all perfect common sense economics and does not need further explanation. However, this view works perfectly for material goods, but is no longer applicable for producing information, as the economics of information distribution is different from the distribution of material goods. With an advanced and rapidly advancing infrastructure for distributing information the sharing of information is almost without costs. Via Internet (including the future Semantic Web and Grid) the goods that can be delivered are in the first (and almost only) place information. As information may be literally about anything, and the needs for information may grow and change with increasing social complexity, ready made information sources may become scarce. To be precise: the processing and organizing of information is getting too costly in face of the diversification of personal needs. The CIA Fact Book is a well organized resource of information, but almost anyone consulting it needs some selective processing to discover that it may or may not be of any help, and there is no Librarian available to act as a brooker between this resource and the clients needs. Therefore, even if information (data) is well organized, this organization may not fit the dynamics of information needs. The Web is however the example of what is to be expected in future information management. The information is distributed in a totally incoherent way and common sense (shared ontologies) and large computational resources (Grid) may serve this information.

This digression on the special nature of services that produce information is necessary because it puts an emphasis on the fact that the matching between the information needs of a client and information resources needs active brokering or at least a provider that is capable of configuring these resources. It means that the focus, i.e. the crucial phase in

obtaining services is in the negotiation and the resulting task distribution.

References

[Breuker, 1994] Joost Breuker. A suite of problem types. In Joost Breuker and Walter Van de Velde, editors, *CommonKADS Library for Expertise Modelling*, pages 57 – 88. IOS-Press/Ohmsha, Amsterdam/Tokyo, 1994.

[Clancey, 1985] W.J. Clancey. Heuristic classification. *Artificial Intelligence*, 27:289–350, 1985.

[Clark, 1996] H.H. Clark. *Using Language*. Cambridge University Press, 1996.

[Golbeck *et al.*, 2003] J. Golbeck, B. Parsia, and J. Hendler. Trust networks and the semantic web. In *Proceedings of Conference on Cooperative Intelligent Agents 2003*, Helsinki, F, 2003.

[Grice, 1975] H. P. Grice. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and semantics: speech acts*, New York, 1975. Academic Press.

[Laurillard, 2002] D. Laurillard. *Rethinking University Teaching*. Routledge, London, 2nd edition, 2002.

[Levelt, 1992] W.J.M. Levelt. *Speaking: from intention to articulation*. MIT-Press, 1992.

[Minsky, 1985] M. Minsky. A framework for representing knowledge. In R. J. Brachman H. J. Levesque, editor, *Readings in Knowledge Representation*, pages 245–262. Morgan Kaufmannn, 1985. originally published in 1975 in Proceedings of TINLAP-75.

[Pinker, 1994] S. Pinker. *The language instinct*. Harper Collins, 1994.

[Roure *et al.*, 2005] David De Roure, Nicholas Jennings, and Nigel R. Shadbolt. The semantic grid: Past, present and future. *To appear in: Proceedings of the IEEE*, 93(3), 2005.

[Sacks *et al.*, 1974] H. Sacks, E.A. Schlegoff, and G. Jefferson. A simplests systematics for the organization of turn taking in conversation. *Language*, 50:696–735, 1974.