CONFERENCIAS

TOWARDS A REALISTIC DIALOGUE MODEL

Giacomo Ferrari

Dipartimento di Linguistica - Università di Pisa
Via S.Maria 36
56126 PISA - Italy
ferrari@icnucevm.cnuce.cnr.it

Abstract:

In this article a brief sketch of the most relevant approaches to discourse and dialogue modeling is proposed. The following three main ingredients for a dialogue model are identified, a theory of focus and focus shifting, a theory of intentionality, and a theory of discourse (syntactic) structure. All of them are briefly described and the major limitations and inadequacies are also put into light, generally deriving from the fact that research have been carried on single types of discourse. An integrated framework, where all the three ingredients find a place and can be generalised, is finally presented. This is not to be taken as a theoretically foundational article, but as a proposal of general features for a realistic theory of dialogue and discourse.

1. Introduction

The history of Natural Language Understanding by computer and Man-machine Interaction in Natural Language, has often been not only marked by the refinement of formal and computational linguistic theory, but has also seen the evolution of the model of interaction . Early systems were based on the assumption that natural language is the natural way of interaction *par excellence*. However, it was soon clear that the ability of the computer to interpret single sentences and give an answer to each question in isolation has nothing natural, even though natural language is used. Naturalness can be only attained by embedding in manmachine systems those mechanisms that make an interaction a connected flow of utterances. Later on, attempts to introduce conversational elements in man-machine interaction in natural language have been carried, and the notion of dialogue system was substituted to the earlier idea of question-answering; sophisticated experiments have been done, in this field, and different aspects of dialogic communication have been highlighted. The ability of understanding (syntactically) ill-formed utterances, of linking the current utterance to the previous ones, thus limiting the use of full reference expressions, which are often substituted by pronouns or ellipses, the capability of penetrating the meaning of a sentence beyond its superficial interpretation, thus understanding the speaker's unexpressed thoughts, intentions, implications and, even, misconceptions, have been recognised as integral part of the human communication process, and corresponding computational models have been developed and tested (for a detailed account of the history of Man-machine Interaction in Natural Language see **[Ferrari 80, 91]**) . During the first phase, roughly along the seventies, complete running systems have been implemented, and commercial natural language interfaces are in general derivations from those systems. In the second phase, instead, good theoretical results have been obtained, and, sometimes, the contributions of linguistics and philosophic pragmatics have been outrun by the originality of the developed computational models. However, no comprehensive dialogue model have been put together yet from the partial accounts provided by research on dialogue modeling, not to. speak of running systems. This is probably due both to the incredible complexity of the phenomenon of human communication, and to the present inadequacy of the computational tools necessary to the implementation of dialogue models, such as knowledge representation and planning. Despite this lack of organic achievements on the side~-- of computational dialogue modeling, in these very last years a new research trend on multimedial interaction has been launched. The idea of using different media of communication, such as natural language, images, animations etc., all together, besides being a natural extension of a man-machine dialogue concept, has been, probably, suggested first by the increasing fortune of

hypertexts. In this field, however, no serious (theoretical) progress have been done, apart from raising the problem. On the other side, very few projects have tackled the problem of multimediality with a sound theoretical perspective, looking at the multimodality of human communication (see [**Reilly et al. 88, 89**]). In the following pages the basic ingredients of dialogue modeling, as they can be abstracted from the existing bibliography, will be examined, also in the light of multimodal communication. Ideas for an integrated dialogue model will be discussed.

2. Ingredients for dialogue modeling

In [**Grosz & Sidner 86**] a complex view of dialogue modeling is presented, which is the result of a summing up of previous research in this field. The two building blocks of a dialogue model, which are described in that article, are the *attentional structure* and the *intentional structure*. The attentional structure is a representation of the concepts which form the subject of the discourse. Any utterance in a dialogue mentions concepts which either have already been introduced by previous utterances, or are completely new. Thus, keeping memory of the discourse subject(s), implies, on one hand the identification of the objects evoked by the linguistic forms, and, on the other, the tracing of any change in the set of those objects. The intentional structure, instead, is the structure of goals, and actions to obtain them, which are shared by dialogue participants. It is generally recognized that many dialogues are motivated by one of the participants having some precise goal, to the satisfaction of which (s)he believes his utterance(s) is/are a contribution. On the other side, (an)other dialogue participant(s)'s task is to try to infer the entire goal structure of his dialogue partner . A third aspect of dialogue modeling, the syntactic one, is only alluded to in the quoted article. These three aspects of dialogue have been separatedly studied in depth, and can be considered the basic ingredients of a theory of dialogue. The view proposed in that article is, in our opinion, a simplistic restructuring of previously developed ideas, with a very weak attempt of synthesis. In the following pages, a critical revision of those ideas, together with general principles of a unifying view will be discussed.

3. Attentional structure, focus, and reference

The attentional structure is strictly related to the phenomenon of focusing, focus shifting and reference. It is generally assumed that not only a discourse is always about a specific subject, but single discourse partitions can be identified, which are characterized by the fact that they refer to a very limited set of objects and events, which is a subset of the general knowledge involved in the whole discourse. The knowledge partition which is known to both the speaker and the hearer to be the subject the conversation is about, is the *fo c u s;* focus is a recursive notion, as given a dialogue partition about a specific subject, it is possible to identify partitions of that partition which deal with subjects which are embedded in the subject of the main dialogue partition.

CANDOPR: Speaker Believe Hearer CANDO Act
WANT.PR: Speaker Believe Speaker WANT request-instance
EFFECT: Hearer Believe Speaker Believe Speaker Want ACT

where preconditions are broken into subsets of preconditions, such as *cando* and *want.pr,* which describe standard preconditions by means of which philosophers analyse speech acts. Finally, [**Pollack 86**] takes a completely different view, representing plans in terms of speaker's of hearer's belief. The general idea is that a plan is not a general schema of actions, but a schema which is believed by someone to be a plan to reach a goal. This position has two important differences with respect to the previously described ones. First, it allows speaker and hearer to have (believe) different plans to reach the same goal, including false plans; this allows to deal with mistakes, misconceptions, and plan negotiations. The second advantage is that plan recognition, comparison, evaluation, and negotiation can take place within the frame of a modal logics of belief and intention. This is, in the same time, a disadvantage, as a modal theorem prover is not computationally well defined yet. All these variants of the so-called plan-based approach to pragmatics agree on a general assumption which can be summarized as follows: when people interact with one another, always have a goal in mind, which they are trying to satisfy. The interpretation of linguistic utterances (requests) is not only the identification of the literal meaning of the utterance itself, but also a

reconstruction, by some inference process, of the intentions underlying the utterance. Answers are more often responses to this reconstruction, rather that to the meaning of the request. This mechanism allows the treatment of many aspects of pragmatics of natural language. Another, less positive point of agreement is that they, in fact, deal with goal oriented types of dialogues, where questions and answers are implied, and the questioner really has a plan in the world and the hearer infers such a plan, with a mechanism or the other, and produces an answer to the inferred goal. However, if a more general perspective on discourse and dialogue is taken. it will appear that in many cases no complex plan inference is involved in communication. Story telling or teaching does not implies a goal, but simply the will of communicating something. In this case, then, the hearer has no goal to infer, but simply to listen and understand. In a general view of discourse, then, it is no more possible to have a single model for pragmatics, but it is necessary to assume that there are different kinds of discourses, some of which do not rely explicitely on an intentional structure.

Thus, if two dialogue participants are talking about used cars sellers, the following exchange

A.: Charlie's has very good used cars.

B.: Oh yes... deflnitely, the engines run like new ones!

A.: They are controlled and fixed up before selling. B.: Yes, ...and the price is really honest.

A.: The price....is the right one.

will run about two different focuses, the condition of the car engines and the price (of the used cars). The subject of the higher dialogue unit still being the same, two subsections focus subjects which are part of the subject of the main partition. The passage from one focus to the other is called ***focus shifting.***

3.1. Why focusing is important

Focus shifting has a relevant role in the modeling of two aspects of natural language communication, reference and dialogue control. Natural language has a quantity of different means to refer tom the same object, without using the same literal expression. In the previous example, the *they* in the third line identifies an object which has been mentioned before. The context offers only two ***antecedents,*** the *cars* and the *engines,* but a simple focusing principle says that the most recently mentioned object is • more likely to be the referred one. Another very common referring mechanism is ellipsis, as in

A.: Give the address of the nearest car hire.

B.: It is Fifth Street 156.

A.: The telephone number?

where it is obvious that the third utterance is an abbreviation for *What is the telephone number of the nearest car hire?*

The reconstruction of the entire message is possible only if the previous (linguistic) context is accessible. However, reference is not only a question of anaphora or ellipsis, which in general realize a principle of echonomy of expression, but many other forms can be used to refer to already mentioned objects. The fragments

John saw Mary across the road. He hastened to reach th e woman.

A.: I have two dogs, a schaeferhund and a dobermann B.: Animals do realy make your life complete.

A.: Take a hammer, nails, bolts, and a wrench. B.: I don't know where the toolbox is.
A.: I have a picture of the mountains in a sunny day.
B.: I like them very much; the snow is incredibly glowing.
A.: ....this has been taken in summer time.

represente different cases of complex reference. In the first case, reference to *Mary is* done by using the term which identifies the class to which Mary belongs *(women).* In the second case, B moves up in the hierarchy of classes and makes a statement about *a n i m a l s,* which *is* motivated by the mention of *d o g s, schaeferhunds* and *dobermans.* the third case

does not even introduce a case of coreference, but simply of discourse coherence; *hammer, nails,* etc are known to be tools, and tools traditionally are in *toolboxes*.. Thus, the mention of a toolbox, whose position is unknown, is a coherent continuation of the previous utterance. Finally, the last example introduces the picture of something (the *mountain)* whose main property is mentioned by B. (the *glowing snow)* as a general attribute which must appear in the mentioned picture. A introduces a change in the standard conditions, so that B's reconstruction must be modified. Many other cases of complex and indirect reference have been introduced in the literature [Webber 81, 83], but all of them seem to highlight two main aspects of the phenomenon of reference. The first one is that in discourse, as well as in dialogue, objects are introduced once by means of a name or a direct referring expression. After this fist introduction the repetition of the same noun, or nominal phrase is, in many languages, forbidden or, simply, not recommended. Thus, substitutive expressions are used, ranging from the simple anaphora, up to the use of concepts related to the referred one only on the base- of common knowledge or perception. Conversely, once a simple or complex referring expression is met, its referent (antecedent) is to be searched for in a well delimited partition of the text which is the focus space where the referring expression occurs.

Thus, in the sequence

A.: I would like to teach New Haven tonight.
B.: There is a limousine service from the airport to New Haven.
A.: Where does it(l) leave from?
B.: All the airport terminals have a stop.
A.: .. and I can also buy the ticket at the stop?
B.: It(2) isn't necessary, you can buy it(3) directly from the driver.
A.: How does it(4) take to reach New Haven?
B.: Oh.... it(S) is very fast.

it(l) refers to the object currently in focus, i.e. the *limousine se-rvice; it(2)* and *(3)* are related to the new focus *(buy the ticket),* but (2) stands for the whole event, while (3) stands just for *ticket;* it(4) is simply a grammatical indeterminate subject; it(S) goes back to the previous focus, i.e. the *limousine service.* The shifting of the focus is, then, *limousine service/buy ticket (for the limousine service)/limousine service,* thus going from a more general subject, to its subspecification, to go back to the general subject.

3.2. How to deal with focus shifting

According to how the example runs, the most appropriate structure for the treatment of focus shifting is the stack, as it has been propose in **[Grosz & Sidner 83].** In fact, subjects of introduced by a push in the stack, and, when exhausted, are popped from the stack. Thus, in the example, the stack will contain, at first only *limousine service;* when a new subject is introduced, without exhausting the previous one *(buy the ticket),* a new push occurs. Finally, *buy ticket is* exhausted and popped, leaving *limousine service* again on top of the stack. This is a very efficient structure for even more complex cases of focus shifting. However, it is still quite unclear what are the conditions under which it is possible to establish that a subject is exhausted; this is, in fact, the only condition for popping the current subject, or pushing a new subject over the current one.

Under this point of view, focus shifting can, and has been dealt with in two different ways, which can be labeled as top-down and bottom-up.

The top-down approach has been experimented in highly structured task oriented dialogues (**[Grosz 77]**). It is assumed that, in this type of dialogues, the structure of the task to be cooperatively carried imposes a rigid structure to the dialogue itself, so that the shifting of the focus is predictable from the structure of the task. The treatment of focus described in **[Grosz 77]** is based on a dialogue between an expert and an apprentice involved in the task of assembling an air compressor; the apprentice physically does the job, while the expert advises and guides him. The task is very well structured and the steps are well defined; in addition, the task-oriented dialogue is rich in delimiting expressions like *it's finished now, next step is.., now....* etc. which mark the focus change points. The focus management mechanism proposed in **[Grosz 77]** is based on a knowledge base describing all the steps involved in the task of assembling an air compressor, and also all the objects involved in any step, i.e. a complete model of the task. Utterances are mapped onto this knowledge base, thus identifying the subtask being carried and the objects connected to it. Such a space is taken to be the focus space

within which referring expressions must be solved. Once a subtask is finished, the focus changes even if no surface linguistic element marks the shift.

This approach allows the treatment of sentences like

John saw Mary across the road. He hastened to reach the ***woman*** .

as, in the knowledge base, *woman is* connected to *Mary* as a superconcept; thus, this kind of reference can be dealt with by accessing the superconcept(s) of the antecedent.

To sum up, the top-down approach to focusing and focus shifting consists in opening a window over a knowledge base and assume that all the concepts visible from that window can be used as referring expressions, according to some specific rules.

The bottom-up approach is, instead, a very local one. The basic assumption, in this case, is that dialogue participants, in fact, only perceive sequences of utterances and are able to connect sequentially utterance to utterance, without a general view of what the dialogue is about, till the end of it.

[Sidner 83] presents a mechanism which deals with focus shifting based on local rules for the establishment of local focus on the basis of the syntactic structure of utterances. The basic idea is that from a sentence to the following one, a main focus is presented, with different syntactic means, and also other candidates for future focusing are also presented. In fact, Sidner's model proposes two different data structures, one for the current focus, and one for future potential foci. Focus shifting consists in a process of updating both data structures, according to a set of syntactically based rules. In fact, **[Sidner 83]** formulates a set of rules which take the current sentence as input, and, on the basis of its syntactic form (which is the subject, which is the object-j is the sentence passive or active, etc.) and of the content of the potential foci register, establishes (updates) the current main focus and changes the content of the potential foci register. In both approaches a relevant position is still reserved to the study of the specific expressions marking the passage from one discourse partition to the following one. These expressions are known as ***cue words.***

3.3. Bottom-up focus tracing as a hint for dialogue structure

Both the bottom-up and the top-down approaches highlight an important aspect of the phenomenon of focusing, but they involve two different perspectives on the running discourse. A top-down view can be assumed only by either an external observer who knows since the beginning what the dialogue will be about, or by an expert involved in teaching, as this situation is very close to the previous one. A bottom-up approach, instead, is the only possible viewpoint for someone taking part into a real dialogue, without knowing since the beginning what the subject will be, and how it will be organized.

We can, therefore, assume that a bottom-up approach is closer to the actual view of a human dialogue participant, who cannot have a predefined perspective on the global sense of the dialogue itself. A natural dialogue participant, in fact, will hear utterances from the speaker, and certainly update his attentional memory structure according to rules very similar to those presented in **[Sidner 83]**. However this approach will prove very limited unless the hearer, besides an idea of focus shifting, has also a notion of the function of each utterance and of the structure of the dialogue itself.

The point, here, is that a bottom-up approach to focusing is very important, only if dialogue participants are able to associate focus shifting to a notion of dialogue structure based on a functional definition of the dialogue units within which focusing occurs.

4. Intentions, goals, real world and communication

Since the early eighties, the idea that intentionality is the underlying structure for the pragmatics of communication, and planning is the technical means to deal with intentionality, became a dominant one **[Allen 83, Cohen & Perrault 79, Allen & Litman 86]**.

Different approaches have been presented where classic speech acts, or other kinds of similar utterances are interpreted as equivalent to actions in the performance of a plan.

Planning is a traditional branch of Artificial Intelligence which is involved into the automatic solution of problems. A ***problem*** can be represented as a pair of states of the world, the *current state* and the state which is to be reached. This last is also called the ***problem state.*** Thus, if an agent A is currently at home and must reach the railway station, this situation can be represented as a pair of predicates which describe the states of the world: *at(A, home)*

will represent the current state, while *at(A, station)* will represent the final state, i.e. the solution of the problem, or, more technically, the *goal*. This can be reached through a series of intermediate states, i.e. intermediate goals or ***subgoals***, such as *at(A, bus_station), on(A, bus)*. The description of the states of the world can be even more complex, such as *(at(A, bus stationl), near(bus_stationl, home)}, (on(A, busl), heading(busl, station)}*, etc. States of the world care more often described by sets of predicates. A ***plan is*** a sequence of states of the world which connect the initial state to the final goal.

The passage from one state to the other is produced by ***operators***, which represent ***actions***. Thus, the two states *at(A, home)* and *at(A, bus station)*, can be connected by an action *move(a, o, d)*, which describes the action moving from an origin (o) to a destination (d). This action will require some ***preconditions*** to be true, such as being at(a, o), i.e. the agent must be in origin, and will cause some ***effects***, such as at (a, d), i.e. the agent will be at the destination. Preconditions must be already true in the world in order to make the action applicable; if a precondition is not true, the action cannot be applied. The effects are updates of the state of the world, i.e. predicates which become true in the next state of the world. Thus, given a representation of the action *move* as

head: move (a, o, d)
preconditions: at (a, o)
effects: at (a, d)

if applied to the situation presented in our example, will match the state *at(A, home)*, thus linking *a=A* and *o=home*. The execution of this action, with *d =bus station*, will cause the predicate *at(A, bus station)* to be true in the next state of the world. A simple consistency rule will remove *at(A, home)* as it is inconsistent with the newly added fact.

A ***planner is*** a programme which, given an initial State of the world and a final goal, automatically builds a sequence of states of the world which connect them, by using a set of abstractly defined actions.

In order to reach some specific state of the world, it may be necessary to attain a state of knowledge, besides some physical goals. Thus, preconditions to the action of moving towards the- bus station can be both knowing that there is a bus to the railway station, and to know where the bus station is. In this view, the action *move* can be reformulated as

head:	*move (a, o, d)*
preconditions:	*at (a, o)*
	know(a, location (d))
effects:	*at (a, d)*

The natural way of satisfying the precondition of knowing where the destination is, in caseit is not known since the beginning, is to ask someone else for such a piece of information. This is a simple way of mixing real actions with actions of a linguistic import. *Asking is* a linguistic action which creates the necessary conditions for the execution of other real actions. This is the general assumption underlying what is generally called the plan-based approach to dialogue, i.e. that there are goals, or subgoals, which can be reached totally or partially by using actions which have a linguistic realization. Different theoretical accounts have been derived from this general point of view, a plan oriented approach, a speech acts oriented one, and a belief states based account being the most successfull results of this approach.

[Allen 79, 83] describes an experimental system which, given a question in the domain of train departures information, identifies the real goal of the questioner and produces answers which respond to this goal, often involving more information than what was literally required. Thus, the question

At what time does the train to Montreal leave?

produces the (over)answer

At 12 from platform 3.

because the hearer infers from the question that the speaker has the goal of going to Montreal by train, and provides the necessary information to board on the right train, i.e. the departure time, which was explicitely requested, and the number of platform, which is also necessary to know in order to be able to completely reach the initiated goal. The key point in this approach is that the system be able to infer that

- if the speaker asks at what time the train to Montreal leaves, the speaker wants to know at what time the train to Montreal leaves;

- knowing at what time the train to Montreal leaves is a necessary precondition to board on it;
- if someone wants to have the information necessary to reach a goal, it is reasonable to infer that he wants to reach that goal;
- if the hearer has identified a speaker's goal and knows specific facts (the platform number) that, even if not explicitely requested, must be known in order to reach the speaker's goal, he will pass to the speaker also that kind of information.

These inference rules were used, in Allen's system, to connect linguistic acts, such as requesting, to well defined plans, which are assumed to be known to both hearer and speaker.

Later on **[Allen & Litman 86]** propose to realize the connection between linguistic acts and a plan using meta-actions,such as *start-plan, seek-parameters* etc., most of which can be realized in terms of utterances (questions). The difference, and, probably, the progress, with respect to the previous approach consists in treating linguistic acts and actions in the world within the same formalism.

[Cohen & Perrault 79] have also tried to provide a uniform representation for actions and linguistic actions. Their approach consisted in using traditional speech acts (**[Searle 69]**) as a model for linguisic actions. This made necessary to establish a precise representation schema into which speech acts could be represented. Thus a REQUEST speech act is rendered as

REQUEST(Speaker, Hearer, Act)

5. Dialogue syntactic categories

Although not much work has been devoted to this approach, some accounts of dialogue and discourse in terms of syntactic and semantic structures have been provided **[Reichman 78, Polanyi & Scha 84]**. This approach generally consists in establishing a set of labels which apply to discourse segments, according to different criteria.

Thus, **[Reichman 78]** devides a discourse into ***context spaces,*** which are identified by a tuple, describing the *type*, the *goal*, the *speakers*, and other properties. The *type* properties are quite close to high level labels and are organised in a hierarchical order; *issue context spaces* can be *debative* or *non debative* etc., while a *non-issue* space may be a *comment*, a *narrative*, or a *support* etc. The passage from a context space to the other is marked by special units, the ***conversational moves,*** which can be *support, restatement, interruption, return, challenges, prior logical abstraction* etc. All the labels are assigned completely informal descriptions in terms of discourse function, goal, and, sometimes, other characteristics.

[Polanyi & Scha 84] present, instead, a hierarchy of discourse constituents. At the lowest level ***clauses*** **and** ***operators*** are classified, where *clauses* are syntactic units, and *operators* are connectors like *and, or, because, well, so, incidentally* etc.. At a higher level ***discourse constituent units (dcu)*** are recognised, the main constituents of any discourse. These can be *sequential structures* as *lists, narratives, topic chaining* etc., or *expansions*, like *elaborations, topic-dominant chaning* etc. A special kind of *dcu's* are the *adjacency structures*, which involve speaker change.At the top level there are the ***discourse units,*** i.e. socially aknowledged units, which have a recognizable purpose and a specific syntactic/semantic structure. This account is allegedly based on the syntactic/semantic characterization of discourse units, but no clear definition of what the semantics of discourse units must be is unclear. However it relies on a relatively smooth passage from the clause (traditional syntactic) level to the discourse level.

Other computational accounts are those presented in **[Wachtel 86]**, where the complex system of features is used to describe dialogue units, or **[Ferrari & Reilly 86],** where a fourlevel hierarchy of functional labels is used to build a real formal grammar for dialogue.

All the above sketched approaches, as well as possible other discourse description systems have few points persistently common. All of them end up with a discourse representation in terms of a graph which catches sequences and alternatives of dialogue segments. Furtherly, all of them tend to present a formal representation which, in fact is not formal, as relies on a difuse characterization of the function and the goal of any segment.

The most relevant coincidence is a philosophical point. The description of discourse in terms of a discourse grammar is possible only if an external view is taken, as it is the case for formal syntax. In other words, as for sentential syntax, the generative approach relies on the implicite assumption that formal characterization applies to complete constituents, wherever they

are observable, also in the case of discourse the characterization of discourse segments (be they spaces, units, or whatever else) applies to complete segments. However, while it is perceptually credible that the hearer starts the process of structuring a sentence *at the end* of its perception, it is also perceptually credible that there are large classes of discourses, like various kinds of dialogues, where participants are in *the middle* of perceiving a unit, but cannot wait the end of it to show a reaction. The conclusion is that there are many cases of discourse, where the generative approach does not apply, unless it can be used to explain the expectation of the participants into the communication process.

6. Yet another dialogue model model

In the previous sections the main ingredients of a (computational) theory of discourse have been discussed. Despite the criticisms which have been proposed to the various approaches, the importance of all of them has never been put in doubt. The problem is, now, to reconstruct a theory which makes use of these ingredients in a new integrated frame, and the - first step will be to define new requirements for the old pieces of the theory .

6.1. New requirements for traditional ingredients

The three elements which are to be integrated are, then, a theory of focus, focus shifting and discourse control, a theory of discourse syntactic/semantic structure, and a theory of the intentional motivations of communication.

The notion of focus and focus shifting has been used to give a basis to the process of reference resolution, i.e. to identify the referents to which future referring expressions may refer. To this end, it has been observed, in § 3.3., that the most natural way of accounting for this phenomenon is the bottom-up approach, as it is closer to the perception of discourse participants. However, this approach is too simplistic, as it relies on simple syntactic rules, connecting one sentence to the following. The right approach, then, is a compromise between the bottom-up and the top-down approach, where a hearer perceives references in a bottom-up way, but process them in a top-down style, i.e. embeds them in a knowledge structure. In a dialogue, focus shifting must be also traced in such a way as to account for different focus spaces for speaker and hearer, as, in some cases, focusing is matter for a negotiation and agreement between dialogue participants. The notion of discourse structure has been used only to capture the most relevant regularities, which make a discourse or a dialogue coherent. Even in **[Grosz & Sidner 86]**, the need for such a structure is stated without any serious motivation. A theory of discourse and, above all, dialogue structure, is, instead, usefull if designed from the point o view of dialogue participants, i.e. as the motivation for a series of expectation which any utterance raises in both the hearer and the speaker. Thus, if such a theory should state that in defining some object, a plain definition is followed by the mention of an instance, or an example of use etc., this must correspond to a discourse structure as far as a plain definition creates, in the hearer, the expectation for one of the possible continuations. This requirement imposes restrictions on the way how discourse categories are defined; in fact, the categories by which discourse segments are labelled shall match functional characterizations which must be directly perceivable by the discourse participants, rather than by an external a *posteriori* observer.

Finally, the intentional structure must be recognized only in those types of discourse where intentions, goals, and plans come into play. That plans can be used, in communication, in a different way than that described in § 4, has already appeared in **[Schank & Abelson 77, Wilensky 78, Hobbs & Agar 81]**, where it is, in general, proposed that narratives can follow the schema of some plan held by the characters of the narrative itself. However, it is easy to imagine situations where communication occurs without any real world plan being in play, such as describing an art masterpiece to someone else. In these cases there is no intentionality, in the above described sense, but for the mere intention to communicate. Nevertheless, also in these cases communication occurs according to a plan, which has communication itself as domain (see also **[Appelt 82]**). It is necessary that a theory of discourse accounts also for this kind of intentionality .

6.2. General schema

The following theoretical schema tries to accommodate all the ingredients in a uniform frame. It looks at communication from the point of view of each participant, who is assumed to

be involved, at any cycle, into a loop of the following three steps of **perception** (of the message), **interpretation** (contextualization of the message and related inferences), and **reaction.** At the end of the loop, in general the roles of speaker and hearer are inverted. During the first phase of perception, the **receiver** of a communication flow perceives a set of facts which, all together, build a message. These can be one or more utterances, gestures faces, or combination of them. He/she is able to assign to each of them a structure and a conventional meaning, which are the result of his/her knowledge of those units being produced according to a communication convention. With this respect, there is no difference between linguistic and non verbal communication. During this phase, if a speaker utters

I want this / {pointing to a book}

the hearer will be aware of a sentence, expressing the will of the speaker to get something, and of a referring gesture. The second phase of interpretation is the more complex one and, in theory, can be split into two subphases, one of **strict interpretation,** and one of **expansion.** Strict interpretation consist in identifying all the referents in context. This operation includes both reference resolution and the identification of new objects. It is strictly related to the semantic interpretation, as it consists in "opening a window" on the knowledge the hearer has about the discourse itself. This is not necessarily a unique operation, as it was in **[Grosz 77]**, but involves different alternatives, concept instantiation being the simplest. In fact, the knowledge about the subject of the discourse may derive from common knowledge, perception or the beliefs the hearer ascribe to the speaker. In this way, focusing is guided by the input (bottom-up), but implies a complex knowledge-based precess where concepts can be instantiated or acquired.

The expansion phase consists in a series of relevant inferences deriving from comparisons and evaluations of the speaker's and hearer's knowledge about the discourse subjects. These include the attribution of a *discourse act type,* by inference from the (syntactic) type of the message and the state of knowledge and belief of the hearer, the inference of the speaker's intentions, the detection of misconceptions and misunders-tandings, and other relevant facts about the message.

The third phase, reaction, consists in the building of a response to the stimulus. This may involve the reconstruction of and cooperation with the speaker's plan, but is not limited to it. There are other situations where the response requires a physical act, or the transmission of knowledge. In all cases, besides the possibility of meeting a real world plan, also the choice of a way how to build the response is necessary. It is, therefore, convenient to distinguish between *substantial* (real world) plans, and *communication* plans, which are about the way of communication.

6.3. What is where

At this point it is necessary to locate the position and the function of the ingredients described above; in other words where, in the framework presented at § 6.2., are involved the three components of traditional theories.

6.3.1. Focus

Focusing is viewed as a general phenomenon of anchoring **acts of reference,** be they verbal or non verbal, to pieces of knowledge representing a model of the world related to the discourse itself. This complex characterization tries to accommodate together all the levels of knowledge involved in the comprehension and interpretation of a message. In fact, reference may involve *common knowledge,* like

I saw a dog barking in the street

which requires the linking of the noun phrase *a dog* to a generic concept of DOG, if it is present in our knowledge base, *referential knowledge,* like

Look ..... how big is that dog!

where the linking involves a specific instance of DOG, well identified in the real world. Also *perceptual knowledge* can be involved, as in

Give me that one

where anchoring requires the correct identification of the pointed at object, its recognition, and the identification of the right corresponding concept. Finally, the building of *new knowledge,* via the acquisition of the speaker's beliefs, can occur, as in

Can you sell me a bridge bolt like this?

where, even if the term *bridge bolt is* unknown to the hearer, he/she can build a specific piece of knowledge where it is stated that "the speaker believes that an object of the presented shape is a bridge bolt, and probably it is".

Focusing is dealt with according to two structures, a stack, which provides a list of recency, and the hierarchical structure provided by the knowledge base itself.

A still open research area is about the possibility of accessing the stack according to some heuristics.

6.3.2. Dialogue categorization

A syntactic theory of discourse and dialogue acquires new importance in this frame. This relies on two levels of description, a perceptual syntactic and a discourse-functional one (see also **[Reilly et al. 88, 89]**). The first level describes **communication acts,** i.e. any act, verbal or non-verbal, which carries a non contextual meaning. They are described as structured units, endowed with meaning. Thus, utterances can be described by means of their syntactic and semantic structure; pointings can be described as pseudo-lexical items, and faces as statements about the psychological attitude of the dialogue participant.

The second level describes **discourse acts,** i.e. communication units relevant to the discourse, in terms of their *discourse functions*. Discourse acts may consist in one or more communication acts, both temporally cooccuring or sequential. The discourse functions are in all similar to speech acts, as they are defined in terms of felicity conditions, involving the state of knowledge and belief of both speaker and hearer (**[Ferrari to appear]**). Discourse functions belong to the dialogue categorization, but are determined during the subphase of extension.

In this approach discourse act types, or discourse functions, can be used as labels for discourse segments, and form higher units in terms of legal sequences; but in the same time they are defined in terms of attitudinal states of speaker and hearer, thus working as a bridge between discourse syntax and the reconstruction of the intentions and goals of dialogue participants.

6.3.3. Planning

Plans play a double role in the presented framework. There are real world plans, directed to the solution of real world problems, where, sometimes, communication with other agents is necessary in order to exploit cooperation, and there are also plans about the communication strategy. This leads to a distinction between substantial and communicational plans. The different balance between these two characterises different classes of discourses; in goal oriented dialogues real world planning will prevail, while, for instance, in teaching, communicational strategic plans will be the only ones.

Although both these types of plans can be reduced to the same activity, at the present state of knowledge they are represented in different ways. Real world plans rely on the tradition of planning and description of plans in the field of problem solving. For communicational plans, an explicite representation of discourse strategies, in terms of sequences of discourse acts, with alternatives and choice points, will be preferred. It may be possible that further research identifies general operators, and the possibility of automatically expanding plans, also in the field of communicational planning.

7. Conclusions

After a survey of the most common approaches to dialogue modelling, a new integrated framework has been presented. It is not the report of robust theoretical account of discourse and dialogue, but simply the framework into which further research should be embedded. Where it has been possible, directions for future research have been indicated. The major limit of discourse studies, in fact, has been that all the previous approaches have dealt with single types of discourse. The implicite assumption was that the account given for one kind of discourse could be extended to other types of discourse. The general conclusion of this article is that the results of the studies carried till now should be kept, but the study of different kinds of discourse and dialogues will show that they are structurally different, serve different goals and must be accounted for in different ways. Thus, discourse studies must climb a level of

abstraction, in order to provide a more general modelling way, even if relying on the already developed theoretical approaches.

References

[**Allen 79**] J.F.Allen, A *plan-based approach to speech act recognition*, PhD thesis, Toronto 1979.

[**Allen 83**] J.F.Allen, Recognizing intentions from natural language utterances, M.Brady & R.C.Berwick, *Computational Models of Discourse*, MIT Press 1983, pp. 107-166.

[**Allen 87**] J.Allen, *Natural Language Understanding*, Benjamin/Cummings, 1987.

[**Allen & Litman 86**] J.F.Allen & D.J.Litman, Plans, Goals, and Language, *inIEEE Proceedings*, 74-7, G.Ferrari (ed.) Special issue on Natural Language Processing, July 1986., pp. 939-947.

[**Appelt 82**] D.E.Appelt, *Planning Natural Language Utterances to Satisfy Multiple Goals*, Tech.Note 259, SRI International 1982.

[**Cohen & Perralt 79**] P.R.Cohen & C.R.Perrault, Elements of a plan-based theory of speech acts, *Cognitive Science* 3, 1979, pp. 1 77-2 1 2.

[**Ferrari 80**] G.Ferrari, Interazione uomo-macchina in linguaggio naturale, sviluppi recenti e prospettive, *Atti del Congresso Annuale AICA 80*, Pavia, 1980, pp. 937-950.

[**Ferrari 91**] G.Ferrari, *Introduzione al Natural Language Processing*, Bologna 1991.

[**Ferrari to appear**] G. Ferrari, Speech Acts e Modelli Computazionali del Discorso Interazioni tra Linguistica, Filosofia e Intelligenza Artificiale, to be published in the *Proceedings of the XXIV Meeting of the Societa' di Linguistica Italiana*, Milan 1990.

[**Ferrari & Reilly 86**] G.Ferrari & R.Reilly, A two-level dialogue representation, *Proceedings of COLING86*, Bonn 1986, pp. 42-45.

[**Grosz 77**] B.J.Grosz, The representation and use of focus in a system for understanding dialogs, *Proceedings of UCAI*, 1977, pp. 67-76.

[**Grosz & Sidner 86**] B.J.Grosz & C.Sidner, Attention, Intention, and the Structure of Discourse, *Computational Linguistics* 12, 3, 1986.

[**Hobbs & Agar 81**] J.R.Hobbs & M.Agar, Text plans and world plans in natural discourse, *Proceedings of the Seventh International Joint Conerence on Artificial Intelligence*, Vancouver 198 1, pp. 190- 196.

[**Polanyi & Scha 84**] L.Polanyi & R.Scha, A syntactic approach to discourse semantics, in *Proceedings of Coling84*, Stanford 1984, pp. 4 1 3-4 19.

[**Pollack 86**] M.Pollack, A model of plan inference that distinguishes between the beliefs of actors and observers, *Proceedings of the 24th Annual Meeting of the ACL*, 1986, pp. 207-2 14.

[**Reichman 78**] R.Reichman, Conversational Coherency, *Cognitive Science*, 2, 4 1978, pp.283-328.

[**Reilly et al. 88**] R.Reilly, I. Prodanof, G. Ferrari, Framework for a model of dialogue, *COLING-Budapest*, 1988, pp. 540-543.

[Reilly et al. 89] R.Reilly, I.Prodanof, G.Ferrari, M.Marino, A.Saffiotti, E.MacAogain, N.Sheehy, CFID: a robust man-machine interface system, *Atti del Primo Congresso della Associazione Italiana per l'Intelligenza Artificiale,* Trento 1989, pp. 78-86.

[Schank & Abelson 77] R.C.Schank, R.P.Abelson, *Script, Plans, Goals, and Understanding,* Lawrence Erlbaum, 1977.

[Searle 69] J.R.Searle, *Speech Acts, an Essay in the Philosophy of Language,* Cambridge Univ. Press, 1969.

[Sidner 83] C.Sidner, Focusing in the comprehension of definite anaphora, M.Brady & R.C.Berwick (edd.), *Computational Models of Discourse,* MIT Press 1983, pp. 267-330.

[Wachtel 86] T.Wachtel, Pragmatic sensitivity in NL interfaces and the structure of conversation, *Proceedings of COLING86,* Bonn 1986, pp. 35-41.

[Webber 81] B.L.Webber, Discourse model synthesis: preliminaries to reference, A.K.Joshi, B.L.Webber, I.A.Sag (edd.) *Elements of Discourse understanding,* Cambridge University Press, 198 1 .

[Webber 83] B.L.Webber, So what can we talk about now?, M.Brady & R.C.Berwick (edd.), *Computational Models of Discourse,* MIT Press 1984, pp. 331-371.

[Wilensky 78] R.Wilensky, *Understanding Goal-Based Stories,* Research Report #140, Yale University, 1978.