ISSN: 2036-6728
RIFL rivista italiana filosofia del linguaggio
Natura umana e linguaggio, 1/2009

A naturalistic perspective on dialogue systems

Duilio D'Alfonso
Università della Calabria
dalfi_duk@hotmail.com

1 Dialogue Systems

A *dialog system* is a computer system intended to converse with a human, in a coherent and meaningful manner. Dialog systems have employed text, speech, graphics, haptics, gestures and other modes for communication on both the input and output channel. In recent years, spoken dialogue systems technology has been developed also under the pressure of commercial purposes, and many applications have appeared on the scene, many of which with specific commercial ends.

In short, the two main courses in Dialogue Systems theory and technology are:

- Dialogue Systems endowed with (some sort of) dialogue management
- Conversational simulators, based on pattern-matching technique.

The latter are of limited theoretical interest.

1.1 Restricted domains

Dialogue Systems are almost always the result of a trade-off. Often they supply their limited linguistic knowledge with various strategies, related to the purposes, tasks and domains of application of the system. These strategies are generally inspired by an awareness: the more constraints and expectations the system can impose on a dialogue, the more these constraints and expectations can be used to simplify the interpretation of user utterances.

Thus, in restricted domain such as travel-planning, system designers naturally tend to employ simple linguistic and dialogue processing procedures. Domain-specific keyword/phrase spotting and slot-filling techniques are prevailing methods for utterances interpretation (McTear 2004).

At dialogue level, systems tend to keep the dialogue initiative, and dialogues are inevitably system-driven. The following is a simple example of a dialogue system based on *semantic grammars*:

<deliverymethod> = [<beginning>] <type> [*any] [<end>]
<beginning> = [PLEASE] USE | I WOULD LIKE | I NEED
<type> = standard \{std-shipping\} | express \{exp-shipping\}
<end> = PLEASE | IS OK

for recognizing user input like: "(I would like/I need) standard/express (...) (please)", and behaving appropriately.

Domain-specific dialogue systems, such as travel-planning, play, in the field of human techno-science, the same role as *Drosophila* in genetics: a model organism, simplified and tractable. For instance, allowing only strictly aimed dialogue you can work around the very intricate problem of language generation. Understanding user utterances serves only to (and is designed for) determine the current progress in the dialogue plan or in the planned task to be accomplish. And this current status in task performing (plus the status of dialogue obligation and turn, in a bit more sophisticated systems), completely forces the system's next move.

1.2 The dialogue states

Nevertheless, here it seems to emerge a theoretical relevant point: Participants, in a coherent dialogue, normally make their decisions about the moves to be done in relation to the current status of the ongoing dialogue action.

Roughly speaking, dialogue moves, in dialogue actions, are function of dialogue states. This means that:

- formalizing the representation of dialogue states,
- formalizing the relation between dialogue states and dialogue moves

are two compelling commitment in outlining a logic for dialogue. (Maybe *the* two compelling commitment!)

Focusing on dialogue states, some features of them are worth to note. The state of a dialogue:

- is public and constantly available for every participants engaged in dialogue,
- is cooperatively built up and cooperatively updated,
- is clearly distinct from private knowledge of participants (participants being fully aware of the distinction),
- must be used by dialogue agents as the relevant context of the conversation.

We can imagine the state of a dialogue as a box where participants put some things, and where they can always give a look.

Postponing the question of the linguistic content of the *dialogue box*, it should contain the shared beliefs that are formed during dialogue exchange, up to the current utterance, and not what is logical derivable from that beliefs. It is the system (and the users) that should be equipped with an inference engine.

We could illustrate the point with the following excerpt of a collaborative problem solving interaction, between system and user, in the Circuit-Fix-It Shop system (a prototype of problem solving application of dialogue system).

1 S: Is there a wire between connector 84 and 99?
2 U: It is not there.
3 S: Add a wire between connector 84 and 99.
4 U: I need help.
5 S: Obtain a wire.
6 U: Okay.
7 S: Connect metal of one free end of the wire ..

You can note the cumulative and cooperative process of the common ground construction, and trace the transfer of private knowledge into the dialogue box, and vice versa. Remarkable also the interaction between private and public knowledge in determining participant's moves.

Some systems have achieved a certain "flexibility", appearing rather natural, implementing a user's model, i. e. involving AI technique of planning, an inference engine and some sort of representation of user's beliefs, desires and intentions on

which to make inferences (this is the so-called Beliefs Desires Intentions (BDI) approach to conversational agency). So, if you ask such a system for a departure time, probably it tells you the departure gate, too!

U: When does the train to Barcelona leave?
S: 3.15 at gate 3.

Here some Action-Effect rule is involved, that allows to go beyond the literal interpretation of the user's utterance.

1.3 Dialogue as a rational activity

Is Dialogue System theory and technology a chapter of the study of the human cognition and, ultimately, of the human nature? Answering to this question is not a theoretical matter, an analytical problem that we can solve a priori (reasoning about armchair knowledge). This is an empirical issue, that should be decided empirically. But we can advance some claims to dialogue system technology.

We have to intend dialogue as a purposeful activity in a broad sense: dialogue as rational activity, in which we are involved as rational agents. The more basic purpose in engaging dialogue is mutual understanding: I want you to believe that *p*. So I could involve you in a conversation whose topic is *p*. But in the course of the dialogue, I could conceive supplementary purposes: if I know that *q* explains *p*, I could tell you *q*, as an explanation of *p*. Alternatively, you could ask me an explanation of *p*. The dialogue might proceed entering in a subtopic whose subject is *q*. Naturally, each of us could introduce a new topic. And so on, recursively. Surely, dialogue is a nondeterministic process. The prosecution of a dialogue is not a single-valued function of some finite set of previous dialogue states (at least, it is a multi-valued function of them). And this is the main difficulty in modelling dialogue. Nevertheless, dialogues continue to show some degree of "structure". Even if a conversation is domain-independent, even if there is not a clear underlying purpose, it unfolds

showing some structured features.

In other words, modelling dialogue as rational activity involves implementing discourse cohesion and coherence devices. The internal articulation of a topic, the opening of subtopics, the interruptions and recovering of topics in dialogue are essentially the same that in discourse. We can open a window on the "branching processes" in dialogue by implementing the theory of the discourse rhetorical structure (discourse grammar). *The dialogue structure is the discourse structure.* To sum up until now, the following figure shows the minimal requirements for a *naturalistic dialogue system:*

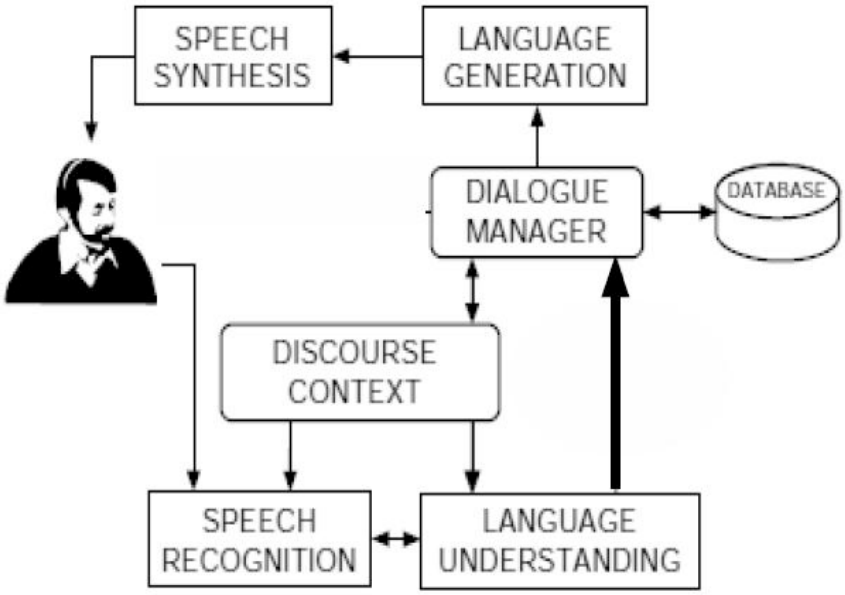


Fig.1: The simplified modular design of a dialogue processor.

ISSN: 2036-6728
RIFL rivista italiana filosofia del linguaggio
Natura umana e linguaggio, 1/2009

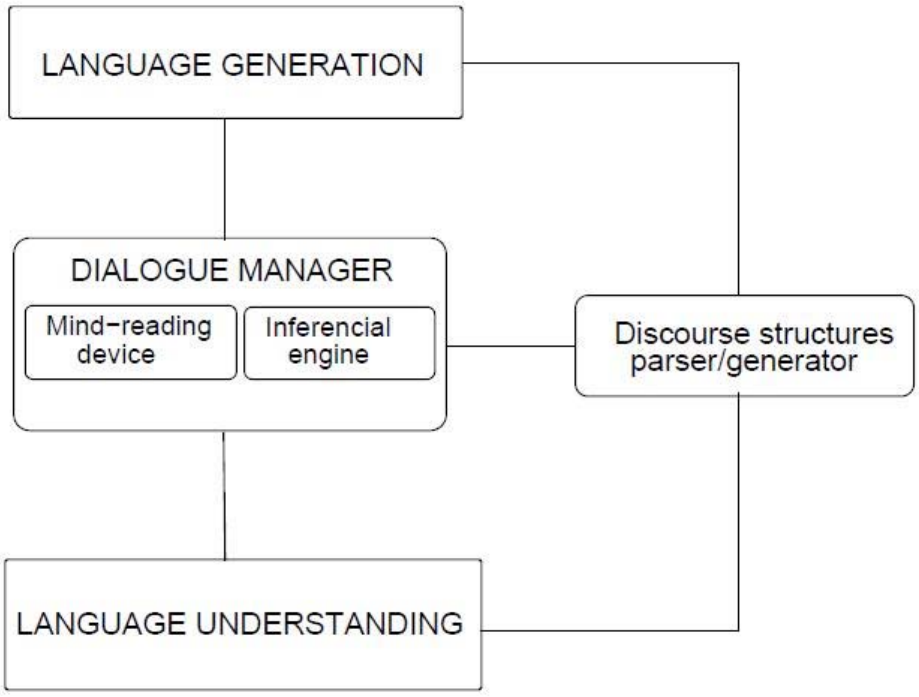


Fig. 2: The core of the dialogue manager.

2 Dialogue Structure

2.1 Syntax and Dialogue

How to "implement" discourse structure analyzer/generator in a dialogue system? If the more basilar function of dialogue is the grounding of propositional content among communicative agents, as a good approximation, you can conceive dialogue as an interactive process of building up discourse tree structures, whose leaves are propositions.

Now, I argue that this process can be formally characterized within the theoretical framework of the *Dynamic Syntax* (DS) (Kempson et al. 2001), augmented with some features of the *Linguistic Discourse Model* (Polanyi et al. 2003).

The fundamental reason in favour of this hypothesis is that, in the Dynamic Syntax approach, either the processing of single sentences or that of "dialogue units" end up

in the same output: propositions. In fact, Dynamic Syntax conceives *grammar* as a set of available strategies for recovering/grounding propositional content from speech interactions in context.

It follows some features of DS:

- Grammar is conceived as a set of procedures for real-time parsing in context;
- Languages syntax is a set of constraints on how context-dependent interpretation is built up on-line;
- Underspecification/enrichment treatment is part of the parsing procedure;
- Production is explained in terms of the same mechanisms as parsing but with an extra filter;
- Context-dependent incremental generation is parasitic on the parsing architecture.

By virtue of these properties, the DS theory of grammar is able to directly handle with *dialogue units* such as:

- Adjacency pairs;
- Cross-speaker ellipsis;
- Shared utterances;
- Speaker/hearer exchange of roles across all syntactic dependencies;
- (Non)-repetitive acknowledgements;
- (Non)-repetitive clarification;
- Repairs.

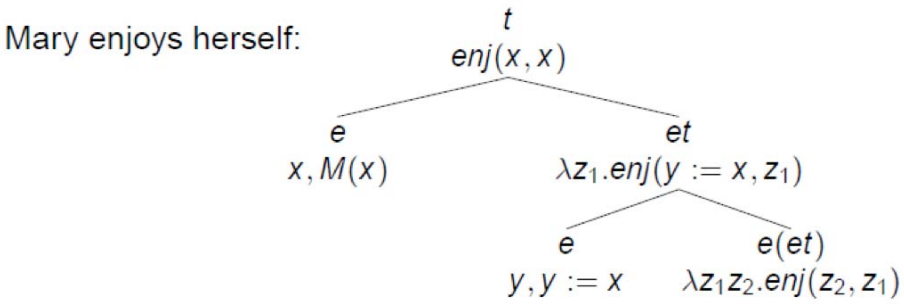
All these dialogue fragments end up in grounding propositional units. And the "discourse tree", in turn, is dynamically built up by composing propositional units.

2.2 The Dynamic Syntax machinery

Instead of theoretically introducing the machinery of Dynamic Syntax, I'm proceeding with an illustrative case.

Since Dynamic Syntax conceives grammar as a set of strategies to recover semantic representation, the following is the semantic representation of the sentence "Mary

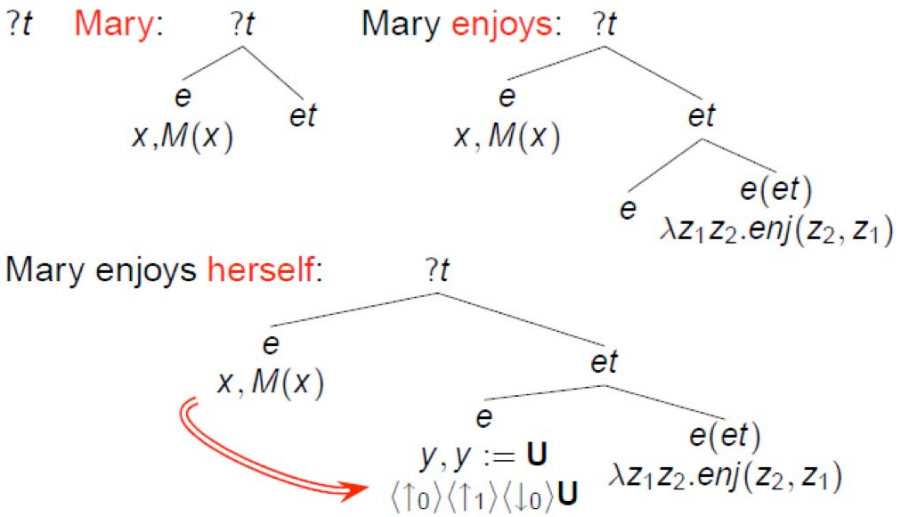
enjoys herself". So, this is the "target tree" that the interpretation process should recover.



Exa

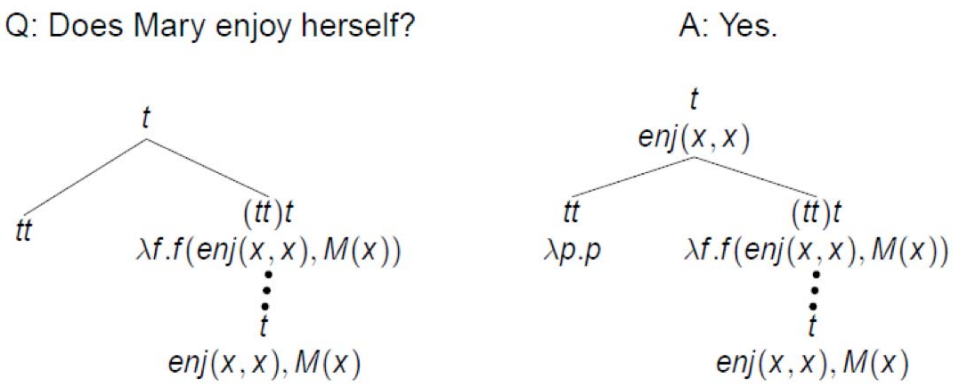
mple 1

In turn, parsing is construed as the process of the on-line monotonic growth of the tree representing meaning, whose starting state is *?t*, that is the requirement of a semantic value of type *t* (i.e., the semantic type of propositions, a truth-value). The process is type-driven and every requirements can trigger subordinate requirements, or sub-goals of the interpretation process. It follows the simplified representation of this process for the sentence "Mary enjoys herself":



Example 2

Let now consider the polar question derived from this sentences: “Does Mary enjoy herself?”. In a “structured meanings” approach to semantics, such as that of Dynamic Syntax, *questions* basically are functions, and their possible answers supply the arguments to these functions, in order to yield a proposition. Thus, every interrogative denotes a function from answer-type denotation to propositions (Aloni et al. (ed.) 2007). Since the answer-type of polar questions, that is the type of “yes” and “no”, is $t \rightarrow t$, than the polar question type will be $(t \rightarrow t) \rightarrow t$. Thus, we can represent as follows the “adjacency pair” (question-answer pair):



Example 3

The previous adjacency pair determines the grounding of a proposition. It may be expanded in a dialogue fragment, in which the grounded proposition is coherently enriched:

Q: Does Mary enjoy herself?
A: Yes, she reads novels, plays tennis,..

Here is clear the emerging of a discourse structure.

ISSN: 2036-6728
RIFL rivista italiana filosofia del linguaggio
Natura umana e linguaggio, 1/2009

2.3 Dialogue and discourse structure

According to the *Linguistic Discourse Model*, there are two fundamental discourse relations: *coordination* and *subordination*. Somewhat simplifying, these two relations rest upon the relations between *theme* and *rheme* of a couple of sentences:

• There is Coordination when the two themes are equal and the two rhemes are different, or vice versa;

• There is Subordination when the two themes (or the two rhemes) are derived from each other.

With these two fundamental relations between pairs of sentences, we can construct three type of basic **discourse parse trees** (X and Y are sentences, FS means Forward-looking Subordination, BS means Backward-Looking Subordination):

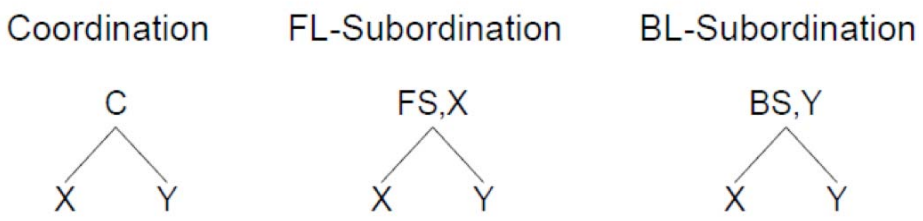
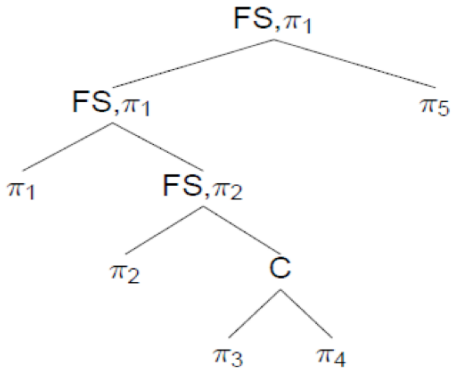


Fig. 3

Note how the root node of subordination inherits the content of the subordinating node. The root node of coordination, instead, should inherit content from both its daughters (something like a *least upper bound*).

Just to illustrate the rhetorical relations at work, here is an example of a discourse fragment, followed by its discourse parse tree:

(π_1) John had a great evening



(π_2) He had a great meal

(π_3) He ate salmon

(π_4) He devoured lots of cheese

(π_5) He then won a lottery

Example 4

The "right frontier constraint" determines what content-node is available for attaching to it the content of the entering sentence, in the ongoing tree construction process (what content is available for subsequent expansion in the discourse). In the same way, this constraint determines the accessibility relation for bounding succeeding anaphora.

It is worth noting that rhetorical relations determining the tree structure of a discourse often involve a surplus of content compared to the truth-conditional meaning of single sentences. In other words, they involve Gricean Cooperation Principle (principle of rationality which governs information exchange) and the deriving *Conversational Implicatures* (Asher, N. and Lascarides, A., 2003).

The discourse tree grows as well as the sentence tree and in dialogue this growth is the result of the cooperation among participants. The growth of the interpretation of the dialogue fragment:

1 A: Does Mary enjoy herself?

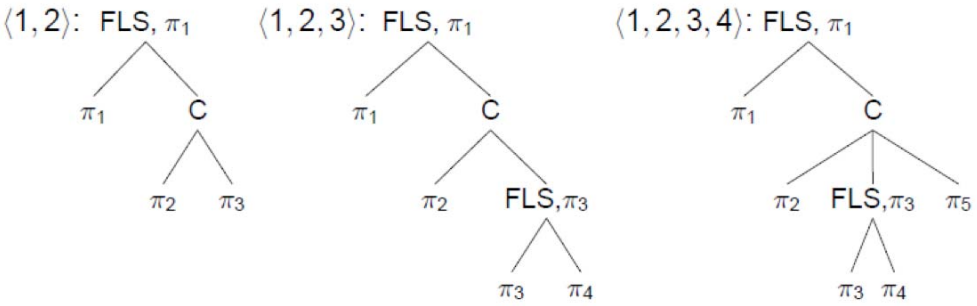
2 B: Yes(π_1) , she reads novels (π_2) , she plays tennis (π_3) ,..

3 A: I like playing tennis, too.

4 B: Really? (π_4) mhm...and she loves shopping (π_5) , too.

may be represented by the following sequence of trees:

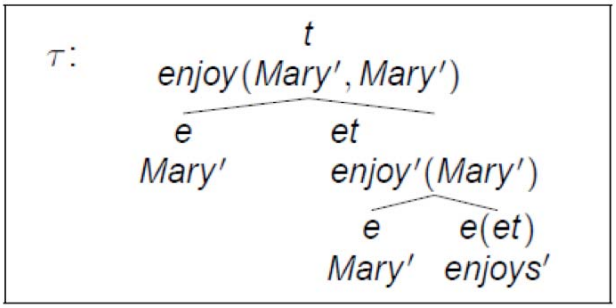
ISSN: 2036-6728
RIFL rivista italiana filosofia del linguaggio
Natura umana e linguaggio, 1/2009



Example 5

Focusing on the first exchange, we can observe how the polar question and the answer "yes" determine the grounding of a propositional content, represented by the semantic tree labelled "τ":

1 A: Does Mary enjoy herself?
2 B: Yes

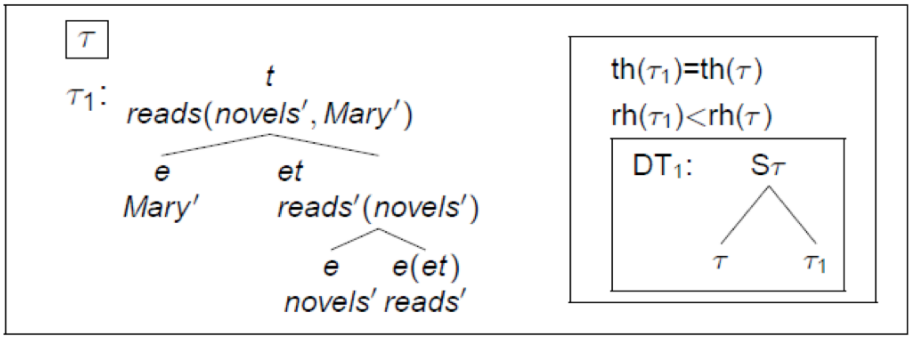


Example 6

Then, once speaker B has uttered "she reads novel" the "box" is enriched by the new proposition, "τ_1", and, the previous proposition working as right frontier, the first Discourse Tree, DT_1, is formed. The relation between the two propositions is of subordination because of the equality of the two themes and the derivation of the rheme of τ_1 from that of τ.

ISSN: 2036-6728
RIFL rivista italiana filosofia del linguaggio
Natura umana e linguaggio, 1/2009

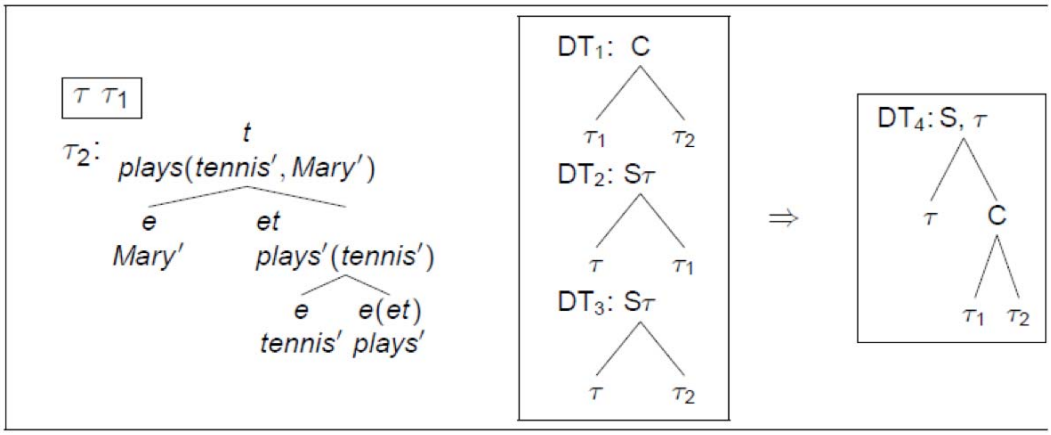
1 A: Does Mary enjoy herself?
2 B: Yes, she reads novels



Example 7

When the speaker B utters his second sentence, "she plays tennis", this utterance, producing the proposition "τ_2", is processed in the context constituted by DT_1. This implies that τ_1 and τ are still available for attaching the entering content, both being collocated at the right frontier of the discourse tree DT_1.

1 A: Does Mary enjoy herself?
2 B: Yes, she reads novels, she plays tennis

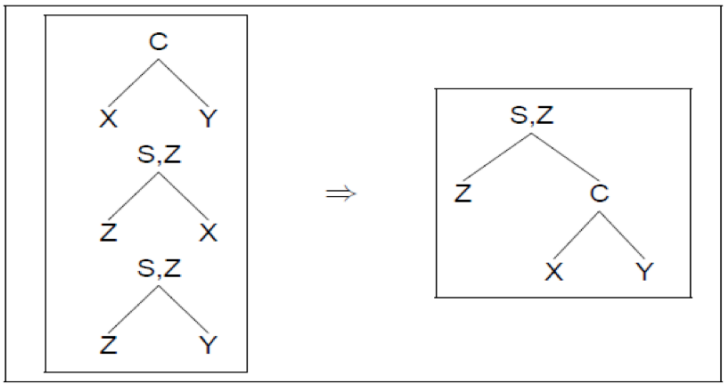


Example 8

Since τ_2 is coordinated to τ_1 and subordinated to τ, this set of relations among propositional units trigger the building up of the discourse tree DT_4, where the three units are integrated in a single discourse parse tree, representing all the rhetorical relations occurring between them.

It becomes evident the need to formally specify a set of tree construction rules. We have to formulate a system of rules at the level of discourse trees, for triggering transitions from sets of discourse trees to single trees.

Just as a minimal example, to give account for the previous case of tree construction process, we should contemplate a rule such the following:



Example 9

In words, this rule establishes that if two sentences X and Y are in the coordination relation between them and are both subordinated to another sentence Z, then a tree in which Z subordinates the coordination between X and Y is to be produced by the system of discourse parsing.

Concluding, what I presented in this paper is a theoretical perspective on dialogue systems: The perspective of integrating discourse rhetorical relations in a constructive theory of dialogue, in the general approach to language processing inspired by Dynamic Syntax. I suggested that along this way we could relax constraints on dialogue imposed by the system, the unique general constraint remaining that of the "rationality" of the conversational agents, in a "gricean" broad sense: agents in dialogue are involved in a cooperative interaction.

Two are the remarkable commitments, in this perspective, both for theoretical interest and complexity:

- the implementation of the dynamic syntax "machinery",
- the formalization of the incremental process of the growth of the discourse parse tree.

References

ALONI, M., BUTLER, A., and DEKKER, P., editors (2007), *Questions in Dynamic Semantics*, Elsevier.

ASHER, N. and LASCARIDES, A. (2003), *Logics of conversation*, Cambridge University Press.

HAREL, D., KOZEN, D., and TIURYN, J. (2000), *Dynamic Logic*, The MIT Press.

KAMP, H. and REYLE, U. (1993), *From Discourse to Logic*, Dordrecht, Kluwer.

KEMPSON, R., MEYER-VIOL, W., and GABBAY, D. (2001), *Dynamic Syntax: The Flow of Language Understanding*, Blackwell, UK.

MCTEAR, M. F. (2004), *Spoken Dialogue Technology*, Springer-Verlag.

PEREGRIN, J. (ed.), (2003), *Meaning: The Dynamic Turn*, Elsevier Science, Oxford, UK.

POLANYI, L., VAN DEN BERG, M., and AHN, D. (2003), «Discourse structure and sentential information structure» in *Journal of Logic, Language and Information*, **12**, 337–350.

RECANATI, F. (2004), *Literal Meaning*, Cambridge, Cambridge University Press.