

# Using Speech Acts in Logic-Based Rhetorical Structuring for Natural Language Generation in Human-Computer Dialogue

Vladimir Popescu<sup>1,2</sup>, Jean Caelen<sup>1</sup>, Corneliu Burileanu<sup>2</sup>

<sup>1</sup>Grenoble Institute of Technology, France

<sup>2</sup>University “Politehnica” of Bucharest, Romania

{vladimir.popescu, jean.caelen}@imag.fr

## Abstract

Usually, human-computer dialogue systems rely on ad-hoc solutions for the component performing speech turn generation, in natural language. However, integration of task-specific and general world knowledge in order to provide a more reliable and natural interaction with humans also through more sophisticated language generation techniques becomes needed. In this paper we present performance improvements of a module simulating in first-order logic Segmented Discourse Representation Theory for language generation in dialogue. These improvements concern reductions in computational costs and enhancements in rhetorical coherence for the discourse structures obtained, and are obtained using speech-act related information for driving rhetorical relations computations.

## 1 Introduction

Most human-computer dialogue systems rely on handcrafted, usually template-based, language generation modules for producing machine’s utterances (McTear, 2002). However, in the last decade or so, with the emergence of research results and ideas from the multi-agent systems domain dialogue systems became more sophisticated, aiming at better responses to user’s requests, via a greater naturalness and relevance of the speech turns produced, in relation to the context of the dialogue and to the users involved (Caelen and Xuereb, 2007), (McTear,

2002). Hence, the natural language generation component itself should aim towards more contextualized and pragmatically situated language productions, involving consideration of rhetorical and actional aspects of language production. In this context, two research trends became distinguishable: (i) a rhetoric-based approach, using formal accounts of discourse originally designed for language interpretation: thus, theories such as Rhetorical Structure Theory or, more recently, Segmented Discourse Representation Theory - SDRT (Asher and Lascarides, 2003) have been adopted for natural language generators (Danlos *et al.*, 2003); (ii) a speech-act based approach, relying on speech act theory (Vanderveken, 1990-1991) or on extensions of it has been used in several systems (Stent, 2002).

In this paper, we show performance improvements for a SDRT-based rhetorical structuring component of a task-oriented spoken dialogue system; these, triggered by the usage of speech acts, consist in: (i) reductions in computational costs involved by discourse structure update, and (ii) improved selection capabilities for choosing the most coherent discourse structure, out of several possibilities.

The paper is structured as follows: the next section provides a brief overview of the baseline rhetorical structuring component, the third one advocates the usage of speech acts in rhetorical structure update, through a discourse update algorithm; then, a discourse update example is presented, allowing comparisons between the baseline approach and the one integrating speech acts; finally, conclusions and pointers to further research are put forward.

## 2 Logic-Based Rhetorical Structuring Component

Our team has designed a rhetorical structuring component integrated in a natural language generation module of a task-oriented spoken dialogue system. In this context, seventeen rhetorical relations have been chosen, in the framework of SDRT, namely:

- first-order rhetorical relations - *Q-Elab*, *IQAP*, *P-Corr* and *P-Elab*, with informal semantics as in (Asher and Lascarides, 2003), that are strongly related to *temporal* aspects in dialogue, hence used in an approximate manner, specific to the type of dialogue concerned (i.e., conversations involving negotiations on time intervals of resource availability);

- second-order rhetorical relations - *Background<sub>q</sub>*, *Elab<sub>q</sub>*, *Narration<sub>q</sub>*, *QAP*, *ACK* and *NEI*, with informal semantics as in (Asher and Lascarides, 2003), that are less constrained by the temporal aspects of the dialogues concerned, hence used in a manner closer to that specified in vanilla SDRT;

- third-order rhetorical relations, specific to monologues and used to relate utterances within a speech turn, generated by one of the speakers (either the human or the machine) - *Alternation*, *Background*, *Consequence*, *Elaboration*, *Narration*, *Contrast* and *Parallel*, with semantics as in vanilla SDRT (Asher and Lascarides, 2003).

Each of these 17 rhetorical relations is expressed as a predicate in first-order logic; each such predicate is expressed in terms of other predicates instantiating actions, operations and relationships between entities. These entities are objects either in a task-independent discourse ontology, or in a task ontology, as described in (Popescu *et al.*, 2007); these predicates take as arguments objects either in the discourse ontology, or in the task ontology (the entities expressing the semantics of the two utterances due to be related via a rhetorical relation. The predicates expressing the semantics of the rhetorical relations are linked through the usual connectors in first-order logic, namely  $\wedge$  (“and”),  $\vee$  (“or”),  $\neg$  or  $\Rightarrow$  (implication); furthermore, each predicate in the discourse ontology is expressed in terms of several predicates in the same ontology and of objects in either of the two ontologies.

## 3 Speech Acts in Rhetorical Structure Computation

Previous studies of our team advocated for the correspondences that exist between pairs of speech acts (Vanderveken, 1990-1991) (customized for human-computer dialogue) and mapping tables have been proposed, using a spoken dialogue corpus, acquired via the Wizard-of-Oz method in the context of a meeting room reservation task (Caelen and Xuereb, 2007).

The taxonomy of speech act *types* proposed by our team supposes that human-computer dialogue is a coordination of actions according to some rules (in order to reach a present or future goal). Hence, the interaction proceeds through an exchange of acts, each one having two components: (i) a propositional content, expressing the semantics conveyed by the utterance produced, and (ii) an illocutionary act that characterizes the utterance in terms of language *use*. Certain acts are *performed* in order to determine changes in the state of things -  $F^A$ : performing an action (denoted by “DO”),  $F^F$ : determining (a speaker) to perform an action (denoted by “MAKE-DO”); other acts are epistemic in nature, that is, they aim at determining changes in the discourse state or mental states of the speakers -  $F^S$ : informing a speaker about certain facts (denoted “MAKE-KNOW”),  $F^{FS}$ : asking (a speaker) about certain facts (denoted “MAKE-DO-KNOW”). Finally, there are two act types that are deontic in nature, i.e., they create obligations (necessities) or give choices (possibilities) -  $F^D$ : compel (a speaker) to do something (denoted “MAKE-MUST”),  $F^P$ : give a speaker choices of doing something (denoted “MAKE-CAN”). Each utterance is characterized by one speech act type, computed, in our architecture, by the dialogue controller for machine turns and by the pragmatic interpreter for user turns (Caelen and Xuereb, 2007); for each pair of utterances one thus has a pair of speech acts and, from a rhetorical point of view, a set of rhetorical relations connecting them.

The point we make here is that the set of rhetorical relations connecting a pair of utterances is conditioned not only by the semantics of the utterances (expressed as logic forms), but also by the speech acts characterizing them from an illocutionary point of view; an extensive corpus study regarding this

$F_U^{FS}$ :	Where can I find book “X”?	
<b>Possible answers of <math>M</math></b>		
$F_M^S$ :	It is at the end of this corridor	<i>QAP</i>
	The plan of the book shelves is down the entrance hall	<i>P-Elab</i>
$F_M^{FS}$ :	Is it for a scientific report you have to write?	<i>Elab<sub>q</sub></i>
$F_M^P$ :	You can take either the hardcover edition, or the DVD edition of this book	<i>P-Elab</i>

Figure 1: Speech acts and rhetorical relations: some examples.

for each utterance  $\alpha$  to be added to the dialogue SDRS:

1. read its corresponding logic form  $K(\alpha)$ , through a query to the *dialogue controller* (Caelen and Xuereb, 2007);
2. for each utterance  $\beta$  already in the dialogue SDRS:
  - (a) read its corresponding logic form  $K(\beta)$ ;
  - (b) **read the pair  $(\gamma_\alpha, \gamma_\beta)$  of speech acts for this utterance and the utterance at step 1.;**
  - (c) **retrieve the set  $P$  of rhetorical relations authorized by the pair of speech acts read at 2.(a);**
  - (d) for each rhetorical relation  $\rho$  in set  $P$ :
    - i. read the semantics  $\Sigma_\rho$  of rhetorical relation  $\rho$ ;
    - ii. compute the truth value  $\gamma$  of the proposition  $\Sigma_\rho(K(\alpha), K(\beta))$ ;
    - iii. if  $\gamma = \text{FALSE}$ , then go to step 2.(c); else add  $\rho$  to the set of rhetorical relations in the SDRS and  $\alpha$  to the set of utterances in the SDRS and go to 2.(c).

Figure 2: Dialogue SDRS updating algorithm.

problem is provided in (Caelen and Xuereb, 2007).

An illustrative example in this respect is given in Figure 1, where we have two speakers, the human subject (denoted by  $U$ ) and the machine (denoted by  $M$ ), and that  $U$  tries to reserve a book in a library.

Using corpus-drawn examples of the type presented in Figure 1, our team has shown that for each pair of speech acts in dialogue, only some (usually, two or three) rhetorical relations (out of all the 17 considered) are *authorized* to connect the utterances involved (Caelen and Xuereb, 2007).

These results are used for refining the set of candidate rhetorical relations in (segmented) discourse structure - SDRS update, according to an improved version of the algorithm presented in (Popescu *et al.*, 2007), by taking into account speech acts in rhetorical structure update.

A rather informal statement of this improved algorithm is presented in Figure 2; steps added in the present version of the algorithm are shown in bold-

face. A rough estimation of the reductions in the computational cost involved by discourse structure updating can be computed supposing that the SDRS to be updated already contains  $N$  utterances, that the total number of possible rhetorical relations between utterances is  $R = 17$ , and that the average number of rhetorical relations authorized by a certain pair of speech acts is  $M$  (usually, 3, according to our studies). Furthermore, assuming that the time needed to read or retrieve logic formulas or speech acts is a negligible constant (since these elements are computed by the dialogue controller, independent of the language generation component (Caelen and Xuereb, 2007)), the computational cost of updating the SDRS with one utterance is  $N \times R$  proofs, since each of the  $R$  rhetorical relations needs to be checked for each of the  $N$  utterances in the dialogue SDRS. We suppose that the time needed to prove a rhetorical relation between two utterances is a constant,  $T$ , thus the computational cost could be evaluated at  $N \times R \times T$  without speech acts, and at  $N \times M \times T$  with speech acts, hence a reduction of  $R/M$  is achieved. For the average values of  $R$  and  $M$ , the computational cost is reduced around 6 times when using speech acts.

#### 4 Discourse Structure Update Example

In order to illustrate the augmentation of the pertinence for an updated SDRS, we consider the dialogue below (here,  $\pi_i$  denotes the label of the  $i$ -th speech turn):

- $U : \pi_1$ : Where can I find some book about “F”?  
 $M : \pi_2$ : You want a book on “F” written by whom?  
 $U : \pi_3$ : What’s available?

From this point on, the machine is supposed to answer that books by authors “A” and “B” are available on the subject “F” and to give the user the opportunity to choose between these two authors; this drives  $M$  to produce two utterances, as an act of informing the user (a  $F^S$ ), and as an act of giving

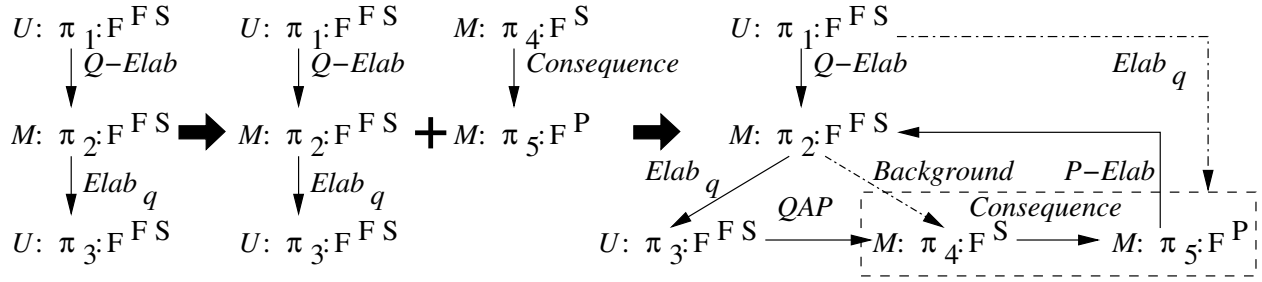


Figure 3: Discourse structure update process.

him a choice (a  $F^P$ ); for these, only logic forms are available (from the dialogue controller (Caelen and Xuereb, 2007)); however, for the ease of comprehension, possible linguistic forms for them are given, in italics, below:

*M : π<sub>4</sub>: We have books by authors “A” and “B”.*

*M : π<sub>5</sub>: Which one you like?*

Then, the machine builds a sub-SDRS using these two utterances, π<sub>4</sub> and π<sub>5</sub>, and adjoins this sub-structure to the dialogue SDRS, formed with the utterances π<sub>1</sub> to π<sub>3</sub>. This process is illustrated in Figure 3; the rhetorical relations between utterances are marked by directed labeled arrows. In da-dotted lines are marked the rhetorical relations computed as valid by the logic-based SDRS update module, but not authorized by the pair of speech acts. Thus, when the machine links π<sub>4</sub> and π<sub>5</sub> through a rhetorical relation, only *Consequence*, authorized by the pair of speech acts  $F^S$  and  $F^P$  in a monologue context, is found between these utterances. Next, the sub-SDRS thus obtained is connected to the dialogue SDRS containing utterances π<sub>1</sub> to π<sub>3</sub> via several rhetorical relations: (i) *QAP* (π<sub>3</sub>, π<sub>4</sub>), (ii) *Background* (π<sub>4</sub>, π<sub>2</sub>), (iii) *P-Elab* (π<sub>2</sub>, π<sub>5</sub>), (iv) *Elab<sub>q</sub>* (π<sub>1</sub>, (π<sub>4</sub>, π<sub>5</sub>)). From these, *Background*(π<sub>4</sub>, π<sub>2</sub>) and *Elab<sub>q</sub>*(π<sub>1</sub>, (π<sub>4</sub>, π<sub>5</sub>)) are not authorized by the pairs of speech acts, which corresponds to our intuitions and to the informal semantics of the rhetorical relations in SDRT (Asher and Lascarides, 2003).

## 5 Conclusions and Further Work

In this paper we have presented several improvements concerning a rhetorical structuring component for language generation in dialogue. These, based on speech act induced constraints, consisted in reduced computational costs for discourse struc-

ture update, and in greater agreement between the discourse structures obtained and human intuitions.

At present, a rhetorical structuring component prototype, integrating constraints induced by speech acts, is under development. In the near future, the discourse structuring module described in this paper will be coupled with other aspects relevant to *spoken* language generation in human-computer dialogue, namely illocutionary force control (Vanderveken, 1990-1991) and (pragmatically-motivated) anaphora generation.

## References

- N. Asher and A. Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.
- J. Caelen and A. Xuereb. 2007. *Interaction et Pragmatique - jeux de dialogue et de langage*. Editions Hermès-Lavoisier, Paris.
- L. Danlos, B. Gaiffé and L. Roussarie. 2003. Document Structuring à la SDRT. *International Workshop on Generation, ACL*, Toulouse.
- M. F. McTear. 2002. Spoken Language Technology: Enabling the Conversational User Interface. *ACM Computer Surveys*, 34(1).
- V. Popescu, J. Caelen and C. Burileanu. 2007. Logic-Based Rhetorical Structuring Component in Natural Language Generation for Human-Computer Dialogue. *Text, Speech and Dialog 2007*, Pilsen, Czech Republic, LNCS, Springer.
- A. Stent. 2002. A Conversation Acts Model for Generating Spoken Dialogue Contributions. *Computer Speech and Language*, 16.
- D. Vanderveken. 1990-1991. *Meaning and Speech Acts*. Cambridge University Press.