Rhetorical Structure for Natural Language Generation in Dialogue

Amy Isard and **Colin Matheson**
Institute for Language, Cognition and Computation
School of Informatics, University of Edinburgh, United Kingdom
{amy.isard,colin.matheson}@ed.ac.uk

Abstract

Many traditional dialogue systems use simple predicates to send information between a Dialogue Manager and a Natural Language Generation system. We propose a flexible RST-style interface to allow for more complex structures and multimodal output, and we place the first stage of content planning under the control of the dialogue management system with access to a system-wide information state.

1 Introduction

In many dialogue systems, a Dialogue Manager (DM) sits at the centre of the interaction, taking user inputs and sending an output specification to a Natural Language Generation (NLG) system. In multimodal systems, there may be a Dialogue and Interaction Manager, and the outputs may go via an intermediate stage where the different modalities are synchronised, and in other cases a planner may take over the duties of dialogue management, but there is usually a need to store the history and to specify the language output. Where the system has a physical component, facial and hand gestures may also be specified.

We propose a flexible interface between the DM and the output modalities which can be used in a variety of human-machine dialogue domains. It is based on RST structures (Mann and Thompson, 1988), and is related to work by (Stent et al., 2004) who used a similar approach in MATCH, their text-based restaurant recommendation system. However in MATCH, the DM sends high-level goals to the text planner, whereas in our systems, the DM performs part of the content selection task (in some cases by communicating with a domain planner and/or knowledge representation module), and sends a structured representation to the NLG system. The potential content to express can come from a wide range of sources, including the dialogue history, the domain knowledge or the task plan. For example, the task plan may describe sequences of actions which need to be carried out in constructing objects (Foster and Matheson, 2008).

This work can be seen in the broader context of attempting to integrate language processing, dialogue management, and NLG more closely.

2 Use Cases

To date, we have used our RST representation in three working systems in varied domains:

- the JAST system (Foster and Matheson, 2008; Foster et al., 2009), which allows a human to collaborate with a robot in building simple wooden toys
- the JAMES system (Petrick and Foster, 2012; Petrick et al., 2012), a robot bartender
- the Beetle II system (Dzikovska et al., 2011), a tutorial dialogue system for basic electricity and electronics

The top level rhetorical structures which we use are the following:

enablement where one situation or action is necessary (but not always sufficient) for another situation to action to occur. e.g. "to build a tower, insert the green bolt through the red cube and screw it into the blue cube" (JAST).

elaboration where one piece of content adds further information about an object which has already been mentioned e.g. "the battery in circuit 5 is in a closed path which does not contain a bulb" (Beetle II).

Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue, September 19–21, 2012, Paris, France.

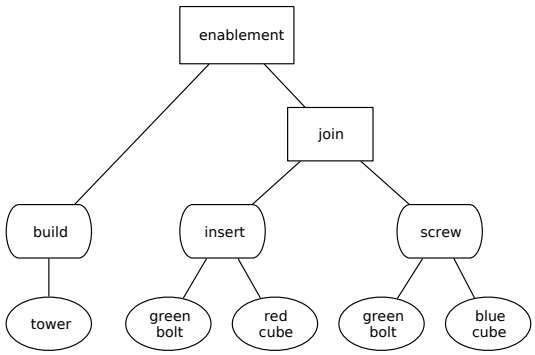

Figure 1: Graph representation of enablement and join relations

```
<output>
 <objects>
  <obj id="o1" type="bolt" color="green"/>
  <obj id="o2" type="cube" color="red"/>
  <obj id="o3" type="cube" color="blue"/>
  <obj id="o4" type="tower"/>
 </objects>
 <rst>
  <relation type="enablement">
   <pred action="build" result="o4"/>
    <relation type="join">
     <pred action="insert">
      <obj idref="o1"/>
      <obj idref="o2"/>
     </pred>
     <pred action="screw">
      <obj idref="o1"/>
      <obj idref="o3"/>
     </pred>
    </relation>
   </pred>
  </relation
 </rst>
 <actions>
  <action type="handover">
   <obj idref="o1"/>
  </action>
 </actions>
</output>
```

Figure 2: Multimodal RST XML for enablement relation

definition where one piece of information defines another e.g. "it means that the battery is damaged" (Beetle II).

join which signifies a simple aggregation of two pieces of content e.g. "hello, what would you like to drink" (JAMES).

A graph of an enablement relation from JAST is shown in figure 1, and a possible surface realisation for this is "to build a tower, insert the green bolt through the red cube and screw it into the blue cube". The multimodal XML representation of the RST is shown in figure 2; as well as giving the content to be spoken, this specifies that the robot should hand over object o1 (the green bolt) to the user.

3 Conclusions

We designed a flexible interface for communications between the Dialogue Manager and the Natural Language Generation components in a dialogue system. We have used the interface in a number of different systems, and shown that it encourages the integration of multimodal output modalities. The systems we have described all use rule-based Dialogue Management or Planning, but the RST could also be used in a statistical dialogue system as long as the NLG component is grammar-based.

References

Myroslava O. Dzikovska, Amy Isard, Peter Bell, Johanna D. Moore, Natalie Steinhauser, and Gwendolyn Campbell. 2011. Beetle ii: an adaptable tutorial dialogue system. In *Proceedings of the 12th Annual SIGdial Meeting on Discourse and Dialogue, demo session*.

Mary Ellen Foster and Colin Matheson. 2008. Following assembly plans in cooperative, task-based human-robot dialogue. In *Proceedings of the 12th Workshop on the Semantics and Pragmatics of Dialogue (Londial 2008)*, London, June.

Mary Ellen Foster, Manuel Giuliani, Amy Isard, Colin Matheson, Jon Oberlander, and Alois Knoll. 2009. Evaluating Description and Reference Strategies in a Cooperative Human-Robot Dialogue System. In *Proceedings of the Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09)*, Pasadena, California, July.

William C. Mann and Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Ronald P. A. Petrick and Mary Ellen Foster. 2012. What would you like to drink? recognising and planning with social states in a robot bartender domain. In *International Workshop on Cognitive Robotics (CogRob 2012) at AAAI 2012*, Toronto, ON, Canada, July.

Ronald P. A. Petrick, Mary Ellen Foster, and Amy Isard. 2012. Social state recognition and knowledge-level planning for human-robot interaction in a bartender domain. In *AAAI 2012 Workshop on Grounding Language for Physical Systems*, Toronto, ON, Canada, July.

Amanda Stent, R. Prasad, and M. Walker. 2004. Trainable sentence planning for complex information presentations in spoken dialog systems. In *Proceedings of ACL*.