Generation Resources and Strategies in the NECA Multimodal Natural Language Generator (MNLG)

Paul Piwek

ITRI, University of Brighton, UK
Paul.Piwek@itri.brighton.ac.uk

1 Introduction

In this paper, we briefly describe the approach to multimodal generation that was taken in the Net-environment for Embodied Emotional Conversational Agents (NECA) project.[1] We outline both the data structures and algorithm that were adopted. It is beyond the scope of this short paper to compare our approach with other approaches to generation, but see [1].

The NECA system is a platform for automated generation of dialogues between embodied conversational agents. The input of the system is a set of facts about an object and an assignment of personality traits and dialogue roles to a team of two interlocutors. On the basis of this information the system produces a computer-animated conversation between the interlocutors about the object in question (e.g., a car in one of the NECA demonstrators).

The modules of the system are arranged in the following pipeline: dialogue planning, multimodal natural language generation, speech synthesis, gesture alignment/timing and media player. Information is passed from one module to the next in the form of an XML compliant Rich Representation Language (RRL) document ([2]). In order to make communication between the modules as transparent as possible, all modules are only allowed to *add* information to the RRL document.

[1] See http://www.ai.univie.ac.at/NECA/. NECA was an EU Project, IST-2000-28580, which ran from 2001 until 2004. I would like to thank my colleagues in the NECA project for the excellent collaboration, in particular, Kees van Deemter, who was the leader of the NECA Team at ITRI.

2 Combining Rules, Templates, Canned Text and Gestures

The NECA Multimodal Natural Language Generator (MNLG; [1]) is provided with its input from the dialogue planning module. This input specifies the dialogue act type, semantics, speaker, addressee, speaker emotion, relation of the current dialogue act to the preceding dialogue act and common ground for the current dialogue act. The output of the MNLG is a set of sentences and gestures that realize the current dialogue act. The temporal ordering of sentences and gestures is specified in terms of a number of relations (*parallel*, *sequence*, etc.). Additionally, the sentences are decorated with their syntactic, semantic and pragmatic analysis. This information is passed on to the speech synthesis module of the NECA system in the RRL document.

The main generation resource of the MNLG is a repository of trees. The nodes of these trees are typed feature structures (implemented using Gregor Erbach's ProFit). Each node has attributes for syntactic, semantic and pragmatic information. Generation starts from an input node –specifying the syntax, semantics (in terms of Hans Kamp's Discourse Representation Structures) and pragmatics of the current dialogue act– and combines the trees from the trees repository, using substitution at non-terminal nodes, for building an output tree with only terminal leaves and with the top node matching the input node.

The current approach subsumes the use of rewrite rules. A rewrite rule can be seen as a tree with only two levels: a mother node and its daugher nodes. Templates and canned text can also be modelled using trees: templates are trees with non-terminal leaves (these correspond with the so-called 'gaps' in a template; in fact, under this interpretation rewrite rules with non-terminals *are* templates), whereas canned text is represented by trees whose leaves are all terminal nodes.[2] Canned text, templates and rules are seamlessly integrated as tree structures in the MNLG. There is no distinction between 'proper' generation resources and 'inferior' templates.[3]

One complication though is that for some purposes it can be useful to

[2]From this perspective, lexical entries for words are a borderline case of canned text

[3]We refrain from calling the MNLG a hybrid system. Reiter's ([3]) definition of such systems seems to presuppose that they integrate different types of operations (proper NLG versus template based operations), whereas our point is that we can deal with both using uniform underlying operations. The distinction between template and proper systems has also been criticized by [4], but with an emphasis on demonstrating that templates and linguistic respectability are *not* mutually exclusive.

have trees which have associated with them computations on the values of features in their nodes. In particular, we use computations which expand part of the semantics in the top node of some trees and instantiate feature values at the daughter nodes with the result. We use this, for example, for mapping discourse referents from the common ground to their identifying descriptions; the computation is known as content determination for referring expression generation. By integrating this computation into the grammar, we depart from the traditional view of a grammar as a self-contained system, especially, if we allow such computations to access non-local information (i.e., information not present in the tree under construction). Also, we allow for gestures as (complex) feature-values on the nodes of our trees.

3 Outlook

The NECA project was completed in May 2004 with the delivery of demonstrators for two domains: the *eShowroom* car-sales demonstrator and the *Socialite* virtual student area demonstrator. The NECA MNLG is currently being re-used in the *Virtual Tour Guide* application for a show-case demonstrator of the EPOCH EU Network of Excellence on Technologies for Open Cultural Heritage. The MNLG is also being packaged for dissemination to the research community. Finally, the XML schemas of the RRL representation language for specifying multimodal agent behaviour are available at http://www.ai.univie.ac.at/NECA/RRL/.

References

[1] P. Piwek. A flexible pragmatics-driven language generator for animated agents. In *Proceedings of EACL (Research Notes)*, Budapest, Hungary, 2003.

[2] P. Piwek, B. Krenn, M. Schroeder, M. Grice, S. Baumann, and H. Pirker. RRL: A rich representation language for the description of agent behaviour in NECA. In *Proceedings of the AAMAS workshop "Embodied conversational agents - let's specify and evaluate them!"*, Bologna, Italy, July 2002.

[3] E. Reiter. NLG vs. templates. In *Proceedings of the Fifth European Workshop on Natural Language Generation*, pages 95–105, Leiden, the Netherlands, 1995.

[4] K. van Deemter, E. Krahmer, and M. Theune. Plan-based vs. template-based NLG: A false opposition? In *Proceedings KI99 Workshop 'May I Speak Freely'*, pages 1–5, Saarbruecken, 1999. DFKI.