Development and Validation of an Architecture for Natural Language Generation (DIVAGALAN)  
TIN2006-14433

Pablo Gervás *  
Universidad Complutense de Madrid

Gabriel Amores †  
Universidad de Sevilla

Abstract

At the start of the project, research in the field of dialogue systems had not addressed issues of natural language generation that are an integral part of the communication cycle. Natural language generation (NLG) research had achieved practical solutions to specific tasks as independent research modules, but they were difficult to interrelate and integrate with other applications. The first goal of the project was the development of a flexible and modular software solution, capable of working with ontologies and the emotional content of messages. This solution should provide a set of reusable software components capable of generating texts suitable for different tasks in different domains. The second goal of the project was to study the application of NLG in spoken dialogue systems in a domotic environment.

Keywords: Multimodal dialogue systems, Ambient Intelligence, Non-verbal communication, Natural Language Generation, Storytelling, User Models, Affective Computing

1 Overall Project Goals

The project proposal contemplated two top level goals: the development of a natural language generation (NLG) application, and its validation in the context of a dialogue system (DS).

The proposal outlined a coordinated project plan aimed at achieving these goals by bringing together the NLG expertise of the NIL research group at Universidad Complutense de Madrid (UCM), and the experience of the Julietta research group of the University of Sevilla (USE) in the development of dialogue systems. Subproject GALANTE (Natural Language Generation for Texts with Emotions) was to concentrate on the task of developing a reusable generic solution for generating textual messages tagged with emotions. Subproject GILDA (Natural Language Generation for Dialogue Systems) was to concentrate on the integration of the generation module in a real dialogue system, and on the task of validating its practical operation in the context of a real application to domotic dialogue systems.

*Email: pgervas@sip.ucm.es
†Email: jgabriel@us.es
An overview of the project chronogram, indicating the tasks originally planned, with their estimated starting and ending times (given in months from the project start), and the institutions participating in each task, is shown in Table 1.

### Table 1: Overview of Project Chronogram

<table>
<thead>
<tr>
<th>N.</th>
<th>Task</th>
<th>UCM</th>
<th>USE</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NLG System Specification</td>
<td>X</td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Dialogue System Specification</td>
<td></td>
<td>X</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Development of the NLG architecture</td>
<td>X</td>
<td></td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>Feasibility study of non-verbal communication in DS</td>
<td></td>
<td>X</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>Evaluation of the NLG Solution</td>
<td>X</td>
<td></td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>Integrating NLG and DS</td>
<td>X</td>
<td>X</td>
<td>16</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>Integration Evaluation</td>
<td>X</td>
<td>X</td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td>8</td>
<td>Extensions to the DS</td>
<td></td>
<td></td>
<td>25</td>
<td>32</td>
</tr>
<tr>
<td>9</td>
<td>Extension of the NLG Architecture</td>
<td>X</td>
<td></td>
<td>25</td>
<td>33</td>
</tr>
<tr>
<td>10</td>
<td>Final Evaluation of the system</td>
<td>X</td>
<td>X</td>
<td>33</td>
<td>36</td>
</tr>
<tr>
<td>11</td>
<td>Management</td>
<td>X</td>
<td>X</td>
<td>1</td>
<td>36</td>
</tr>
</tbody>
</table>

2 Project Achievements

The nature of the achievements of the project is discussed in this section. The ways in which the project has satisfied the goals of the call are discussed in the next section.

2.1 The TAP Architecture for NLG

The main result of the Galante subproject has been the development of TAP (a Text Arranging Pipeline), a software architecture for natural language generation that can be instantiated to develop natural language generation modules tailored for specific purposes. This was addressed by tasks 1, 3 and 5.

The core of the TAP architecture defines natural language generation generic functionality, from an initial conceptual input to surface realization as a string, with intervening stages of content planning and sentence planning. A large number of existing solutions for particular NLG subtasks have been taken into account, as well as the extensive literature on architectures for NLG systems. Existing architectures had focused very specifically on the English language, and adapting them to other languages, such as Spanish, had been identified as problematic. The TAP architecture has been developed and tested over English and Spanish test sets, and the current version shows acceptable coverage over both languages. In this process, several problems arose that had not been considered in the monolingual approach to NLG. The solutions applied to solve some of these problems have given rise to a number of publications [HG06, HCC+07, HG08]. In all cases, engineering solutions based on the use of software patterns have been employed to maximise code modularity and reusability, while minimising code redundancy. All modules have been tested against large data sets, with special attention devoted to efficiency and performance. Results have been satisfactory on all counts.

The TAP architecture has been used as underlying technology in a number of research publications [AHP+07, GPyPSL07, HCC+07]. Solutions based on the TAP architecture have been entered to shared competitive evaluations tasks in the field of natural language generation.
2.2 Extensions of the NLG Architecture

Task 9 of the original plan contemplated the possibility of carrying out a number of extensions to the natural language architecture originally proposed. Achievements along these lines are described below.

A specific subgoal of the project concerned the representation of synthetic emotions and their role in the interpretation of text. In order to provide a solid empirical basis for subsequent developments, a corpus of texts tagged with emotional information was developed. The sentences of the corpus were tagged by human evaluators according to several of the available schemas for representing emotion. The results were analysed to establish an initial model of how humans treat compositionality of emotions over complex sentences [FG06c]. Semi-automatic techniques were employed to design a method for simulating the observed compositional behaviour [FG06a, FG06b]. The resulting information was used to configure a speech synthesizer to produce emotional voice [FGGL07].

Another stated subgoal of the project was to enable the system to handle representations of its input data in terms of description logic ontologies. For this purpose, an ontology for the representation of emotions was developed [FGP07].

A particular subgoal of the original project was concerned with how to allow NLG systems to adapt to the particular communication needs of a given user. This subgoal has been addressed from several points of view: how to build user models and use them to drive the operation of a particular system that communicates information [DG07], and how to evaluate the adaptability of a system to the needs of a given user [DG08].

Another subgoal of the original project was to consider how non-textual data can be fruitfully communicated as text. Early attempts at rendering this kind of data as text quickly identified an important underlying problem: how to convey complex structure within the constraints imposed by text, limited as it is to a linear sequence of sentences. Narrative was adopted as a guiding case study because it reproduced very similar constraints, it is known to play a very important role in human communication, and it also provided examples of interactive application. This effort resulted in publications concerning: story generation [PG06a, GLRM06], story telling [LHG07, LHP07], interactive approaches to narrative [GGP06, PG07, LPN08, CPyPSL07, PCP08], and particular software tools for controlling interactive environments [PN07].

2.3 Non Verbal Communication in Dialogue Systems

Task 4 of the project involved: reviewing existing solutions for the construction of synthetic talking heads, implementing a basic prototype, and integrating it with the DS. This task was fully achieved within the expected timeframe. A virtual butler has been designed and implemented in the MIMUS dialogue system in Seville. A sample screen of the MIMUS system may be seen in Figure 1. MIMUS is a multimodal dialogue system for the control of a smart home. It consists of a set of collaborative agents: a dialogue manager, a knowledge manager, ASR and TTS managers, and a Talking head whose name is Ambrosio.

Ambrosio has been implemented in 3D to allow for more natural and realistic gestures and movements. Its lips are synchronized with Loquendo’s TTS synthesizer. In order to be coherent
with the overall purpose of the system, a virtual character must be invested with the highest degree of credibility: gestures and expressions are almost or quite as important as speech itself. Thus, different behaviour layers have been established, along with conscious (shaking, nodding) and unconscious (blinking, breathing) gestures. In addition, the six basic expressions have been implemented: Happiness, Anger, Sadness, Fear, Disgust, and Surprise. Given that in the current scenario “Disgust” was not deemed to be useful, this expression has been substituted by “Doubt”, to reinforce the human–computer communication, especially given the limitations of state–of–the–art speech recognition nowadays. These expressions suffice to model general facial behaviour, as Figure 2.

2.4 Aggregation Strategies in English and Spanish

Describing the state of different devices in the home domain scenario, where information can be presented and expressed in multiple ways, involves a great complexity for Natural Language Generation systems, and even for human beings. As part of this project, a number of experiments were carried out in order to determine syntactic and lexical aggregation preferences by English and Spanish users. The final goal of this work is to implement aggregation strategies in MIMUS by integrating our system with a more powerful NLG engine developed in the GALANTE project, as explained in section 2.5. Experiments were carried out in both English and Spanish languages. Twenty–four informants were shown 15 print screens of the house in which the devices were in different state configurations. Informants were then asked to describe the state of the devices (12 in Spanish and 12 in English) resulting in a corpus of 180 descriptions for each language. Conclusions of the experiment indicate that ellipsis and coordination are used in both languages, and gapping is additionally used in Spanish. These
results have been employed as requirements for the aggregation module of TAP, currently under development.

The specification of aggregation strategies covers needs arising from specification tasks 1 and 2, but also affects implementation tasks 6, 8 and 9, and evaluation tasks 6 and 10.

2.5 Integration of MIMUS and TAP

Tasks 6 and 7 involved the integration of partial results from both subprojects.

The communication module chosen has been OAA (Open Agent Architecture). This software is a distributed agent communication package that had already been used to connect the internal modules of MIMUS. The main advantages of this framework are described below. First, it is a multiplatform middleware compatible with Linux, Unix and Windows. Second, it provides interfaces to develop agents in Java, C, C++ and Prolog. Third, it includes general purpose agents that help the development and debugging of new agents. Finally, it also includes specific agents that are useful for dialogue applications, like wrappers for third-party ASR and TTS’s.

The integration hereby described substitutes the internal MIMUS NLG engine (APM06) by a more powerful NLG module based on the TAP architecture. OAA is based on “solvable”s, by means of which a server provides a set of services to one or several clients. Whenever the MIMUS Dialogue Manager identifies that some information has to be transmitted to the user, it calls upon a solvable provided by an OAA agent acting as wrapper for the TAP NLG module. This solvable receives an incoming Information State defined by an attribute–value matrix (and an additional parameter defining the language to be used for output) and generates one or more natural language sentences corresponding to it.

2.6 Ontology–based Multimodal Presentation

Task 8 contemplated extensions to the dialogue system beyond the original functionality.
Multimodal interfaces allow for more flexible and natural interactions between human users and computer systems. They benefit from a variety of communication media such as speech, text, gesture, handwriting, etc. As part of the GILDA project, some research has been conducted on providing a Presentation Layer in MIMUS, with a special attention to the Presentation Planner submodule. The Multimodal Presentation layer in multimodal dialogue systems channels all interactions from the system towards the user. As regards its architecture, most multimodal presentation modules are divided into three subprocesses: a Content Planner which determines the information to be offered, and structures such information according to a predefined high-level semantic representation, a Presentation Planner which maps each semantic representation with a specific modality, and a Content Realization module that fleshes out each abstract representation identified in the first module according to the modality selected in the second subprocess. Additionally, the coordination and synchronization of outputs in different modalities is managed at this point.

The Content Planner generates a set of alternative semantic representations. These are then analyzed by the Presentation Planner, which chooses one among them, and, finally, the Content Realization module generates the different outputs according to the chosen presentation modality. In order to choose between the possible presentations, the Presentation Planner makes use of several models: a Context Model, which provides contextual information (e.g. whether there is someone in the room), a User Model, providing user profile information such as whether the user has shown preference for a specific modality of presentation, a Modality Model, which describes the features of each modality channel, as well as the relation between them, and the Dialogue History, which provides relevant information gathered from the dialogue history (e.g. the last modality used). Each of these submodules takes the form of an OWL ontology. At this stage, the formal model presented above has been designed, and some portions of it will be implemented in the remaining months of the project.

3 Indicators of Project Success

Three coordination meetings have taken place so far, two in Sevilla and one in Madrid. The first coordination meeting took place on January 2007 in Sevilla. This was the initial meeting to set up the project, following shortly on the notification that the project had been approved (29/12/2006). The second meeting was held on September 2007 in Sevilla. Results obtained by both subprojects to that point were shared, the architecture for the combined system was proposed, and next steps for both subprojects were discussed. The third coordination meeting took place on March 2008 in Madrid. New functionality obtained by both subprojects was discussed, and next steps were decided. Additional coordination meetings have already been organised for the coming months, to guide the closer collaboration required to bring the project to a successful closure.

The teams of both subproject have been regularly in touch over particular issues such as Eva Florencio’s work (USE) on aggregation strategies in English and Spanish which acted as specification for the aggregation functionality developed for TAP (UCM), and the current work on integrating the TAP generation module with the MIMUS architecture (see 2.5).

With respect to project management, both subprojects can report smooth progress according to the established plan. The delayed communication of the project grant (29/12/2009) had a slight impact on the internal timing of the project. Some of the planned subtasks took less
Table 2: GALANTE Publications

time than had been estimated. Overall, the goals described in the project proposal have either been completed or will be completed in the expected timeframe.

Relevance of the results obtained has been validated by industrial uptake of results from both subprojects: results of the GILDA subproject have been transferred to Indisys (see 3.2) and results of the GALANTE subproject are being transferred to Technosite (see 3.1). The results are original in as much as they arise from the research effort carried out by the two groups involved, and they have given rise to original scientific contributions accepted by the scientific communities in the fields addressed.

3.1 GALANTE

• Scientific and Technological Production.

Scientific production is summarised in Table 2. The various results described in Section 2 have led to a number of publications, involving journals, conferences and workshops.

In addition, the NIL research group hosted the 5th International Joint Workshop on Computational Creativity in Madrid in September 2008. This workshop was partially funded by Acción Complementaria TIN2007-30544-E/ from the Ministerio de Educación y Ciencia.

• Usefulness of Results and Technology Transfer.

The Natural Language Generation modules developed in the GALANTE subproject have been validated by the group’s participation in several shared task of competitive evaluation [HG07, GHL08]. The group’s entries were well placed within the resulting ranking of entries [BG07, GBK08].

In terms of technology transfer, the research group is currently participating in the Cenit Project INREDIS led by TechnoSite, as a research team subcontracted by TechnoSite. One of the tasks of the NIL research group in this project is to transfer NLG technology by developing modules that deal with linguistic presentation in a domotic environment specifically oriented towards people with functional diversity.

• Formation of PhD students and Research Assistants.

Carlos Léon was originally contracted by the project as a researcher. He subsequently joined the project team on receiving a predoctoral grant (Beca Predoctoral UCM), as did Laura Plaza (Beca FPU, Ministerio de Ciencia e Innovación) and Jorge Carrillo de Alborno (Contrato de Personal Investigador de Apoyo de la CAM).
The project has nurtured the development of two PhD theses of project team members (Federico Peinado, 2008; Virginia Francisco, 2008), one DEA (Raquel Hervás, 2006), and 5 MSc theses (Carlos León, 2006; Laura Plaza, 2008; Susana Bautista, 2008; Álvaro Navarro, 2008; Jorge Carrillo, 2008). It is expected that the PhD thesis of Raquel Hervás will also be defended within the time span of the project. All these theses have been supervised or cosupervised by team project members.

- Collaborations with other European and/or International Teams.

Research collaborations have been established with the following foreign research institutions: School of Computing, University of Teesside; Narratology Department, University of Hamburg; Departamento de Engenharia Informática, Universidade de Coimbra, Portugal; Laboratory for Applied Ontology (ISTC-CNR), Trento, Italy; Institute for Creative Technologies (ICT), University of Southern California, Los Angeles, California; Institute for Language Processing (ILP), Athens, Greece; UAM (Universidad Autónoma Metropolitana), Mexico DF, Mexico and UNAM (Universidad Nacional Autónoma de México), Mexico DF, Mexico. These collaborations have involved research stays of team members abroad, visits by foreign researchers, and a number of joint publications [PCP08, GLRMP06, AHP+07, PHGC06, HCC+07, GPyPSL07].

3.2 GILDA

- Scientific and Technological Production.

Four papers have been published so far as part of this project: [FAMP08], [MAPA07], SPF+07, APM07.

In addition, the following paper has been submitted to EACL09 [PAM09], and another one is under production describing the integration between MIMUS and TAP.

- Usefulness of Results and Technology Transfer.

The collaboration between the EPO Intelligent Dialogue Systems (Indisys) and the research group is very close. As a technological leader in Human-Computer Interaction solutions, Indisys is truly interested in the results obtained in the GILDA project. This collaboration has already led to a technology transfer as the pilot study on talking heads carried out in the GILDA project has been incorporated to the avatars used in Indisys. A similar technology transfer is under way with respect to the Multimodal Presentation Framework, suitable for improving presentation of multimedia material by Indisys’ avatars. Both character personification and user cognitive perception issues play a fundamental role in the company’s solutions. Enriching NLG, and therefore dialogue, with personality trait enhancers such as gestures and facial expressions is surely a significant step forward in the development of HCI solutions. Indisys has also subcontracted the Julietta research group to develop human-computer interfaces within the Cenit Project ATLÁNTIDA led by Boeing.

- Formation of PhD students and Research Assistants.

Two postgraduate students have been contracted partially by the project: Ms Eva Florencio Nieto and Mr. Pablo Montoya.
One team member has obtained her PhD thesis (Pilar Manchón, 2009), and Guillermo Pérez is expected to submit his in the coming months. Dr. Gabriel Amores has acted as supervisor or cosupervisor of both.

Mr. Jesús González Marín and Ms. Carmen del Solar Valdés, PhD students at the University of Seville, have also collaborated in the project.

• Collaborations with other European and/or International Teams.

The research group at the University of Seville is currently participating in the submission of two proposals to FP7. Project FUSit (First Use Situation Technologies) aims to leverage the technological and economic potential of multimodal spoken dialogue systems by providing methods that lead to acceptance by their users right from the first contact. It involves as partners: University of Saarland, BMW, CLT, SVox, Nuance, and DFKI. Project E-LAB (EElectronic Assessment of Behavioural competencies) concerns E-assessment, an emerging area of ICT in combination with psychology that aims towards determining human competencies in a more effective way than possible with traditional methods. This project is coordinated by the University of Zurich.

In addition, the research group is participating in the Cenit Project ATLÁNTIDA led by Boeing, as a research team subcontracted by the SME Intelligent Dialogue Systems, a spin–off of the University of Seville, and EPO in the GILDA project. The Julietta research group will develop human–computer interfaces in natural language within this project.

References


