# Dialogue Systems for Virtual Environments

**Luciana Benotti, Paula Estrella, Carlos Areces**
Grupo de Procesamiento de Lenguaje Natural (PLN)
Sección de Ciencias de la Computación
Facultad de Matemática, Astronomía y Física (FaMAF)
Universidad Nacional de Córdoba, Argentina

## Abstract

We present an on-going research project carried out at the Universidad Nacional de Córdoba in Argentina. This project investigates theoretical and practical research questions related to the development of a dialogue system situated in a virtual environment. We describe the PLN research group in which this project is being developed and, in particular, we spell out the areas of expertise of the authors. Moreover, we discuss relevant past, current and future collaborations of the research group.

## 1 Introduction

The goal of this project is to implement a dialogue system which automatically generates instructions in order to help a user to fulfill a given task in a 3D virtual environment. In this context, we will investigate fundamental issues about human-computer interaction. The expected results of the project can be classified in three areas: pragmatics of interaction; information representation and inference; and evaluation of dialogue systems. Once a working prototype is finished, we will adapt it to the specific task of language learning, using the system as a virtual language teacher. Our prototype will teach English to native Spanish speakers. Hence, it will need to understand and produce both languages.

Initially, we will investigate a model of unidirectional linguistic interaction (i.e., linguistic information flows only from the system to the user). In subsequent stages, the model will be extended to allow bidirectional language exchange. For example, the user may ask clarifications to the system or redefine the goal of the interaction.

The architecture of the envisioned dialogue system presents both theoretical and practical challenges. On the theoretical side, heuristics are needed in order to govern decisions such as what to say, when, and how (given the current context). In addition, the system should implement inference methods in order to figure out how to modify the current situation and reach the task goal. The complexity of the theoretical issues is reflected, in practice, in a system of multiple components: a natural language generator, a planner, a 3D interactive environment, to mention a few. Designing and implementing all these components from scratch would require a prohibitive effort. Instead we will adapt tools already implemented and freely available for prototyping this kind of systems, such as the platform *GIVE*[1], *Generating Instructions in Virtual Environments* (Byron et al., 2009).

The quality of each of the components of the system affects the perception users have of it. It is imperative to carry out extensive evaluation. We plan to adapt and apply different evaluation techniques and metrics from the area of Machine Translation to assess the performance of the system.

The plan of the paper is as follows. Section 2 describes the project in detail. Section 3 spells out the expected results as well as their foreseen impact in the Argentinean socio-economic landscape. Section 4 presents the PLN research group including its lines of research. Section 5 discuss past, current and future collaborations that are relevant to the project.

[1] `http://www.give-challenge.org`



*Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*, pages 132–140, Los Angeles, California, June 2010. 

## 2 Description of the Project

This section first introduces the virtual environment in which our dialogue system will be situated, namely the GIVE platform, which is the basic architecture of our dialogue system. Then we explain in detail the tasks that our situated dialogue system will implement, and we spell out the evaluation challenges that such a system poses. We close the section discussing the application of our dialogue system for the task of second language learning.

### 2.1 The Virtual Environment

In the scenario proposed by GIVE (Byron et al., 2009), a human user carries out a "treasure hunt" in a 3D virtual environment and the task of the generation system is to provide real-time, natural language instructions that help the user find the hidden treasure.

In the GIVE setup, the instruction giving system must guide the user through interconnected rooms. The final goal is to get a trophy which is hidden in a safe. In order to achieve this goal, the system instructs the user to perform several subtasks such as deactivating alarms and opening the safe combination by pressing a sequence of buttons on the walls of the rooms.


Figure 1: The user's view of the 3D world

Figure 1 shows a screen-shot of the user's view on the 3D world. On the top of the picture, the current instruction generated by the dialogue system is displayed. The picture shows a closed door and an open door that has an activated alarm (that looks like a red tile) in the doorway. There are five visible buttons in this room (two yellow, two red and one green) and the instruction giver is instructing the user to press a red button. Pressing a button can have different effects such as opening a door, moving an object, deactivating an alarm, etc.

The characteristics of the world, including the functions of the buttons, are described in the world specification by the world designers. The user can move freely around the world (using the direction keys as indicated in the bottom of the screen) but she can loose the game if she triggers an alarm. The user can also ask for help pressing 'H' if she did not manage to read or understand the last instruction.

For the correct definition of the interaction policies of our prototype we need a corpus that provides examples of typical interactions in the domain. GIVE provides tools for collecting such a corpus in the form of a Wizard of Oz platform that records all details of the interaction, thus allowing to easily obtain a corpus of interaction in virtual environments annotated automatically.

### 2.2 The Dialogue System Tasks

From the collected corpus we will begin the design, implementation and testing of our dialogue system. The main components that we will have to design and implement can be organized using the traditional four tasks that a dialogue system should address: (1) content planning, (2) generation of referring expressions, (3) management of the interaction context, and (4) interpretation of user responses.

*(1) Content Planning:* Given the envisioned setup we described before, the first task of the system is to obtain a plan to reach the desired goal, from the current state. The plan will contain physical actions to be performed in the virtual environment. The second step is to decide how to transmit this sequence of actions to the user. E.g, to decide how many actions to communicate per instruction, and how to aggregate them coherently. The result of the action aggregation process can be represented as a tree describing the task structure at different levels of abstraction. The third and final step is to decide how to navigate the tree of actions to verbalize the instructions (for example, post or preorder as explored in (Foster et al., 2009)). We will investigate different

aggregation policies (e.g., aggregating actions that manipulate similar objects) and innovative ways in which to navigate the task tree (e.g., moving to a lower level of abstraction in case of misunderstandings). Plan computation can be solved using classical planners (Kautz and Selman, 1999; Hoffmann and Nebel, 2001; Nau et al., 2004). However, while there are planners that work well when optimized for certain applications, none provides services such as the generation of alternative plans, or the generation of incomplete plans in case of the absence of plan. One of the goals of the project is to design and implement these extensions to classical planning algorithms. We will also study the theoretical behavior (e.g., complexity) of these new algorithms.

*(2) Generation of Referring Expressions:* Once content planning is complete, the next step is to generation adequate referring expressions. This task involves producing a phrase that describes a referable entity so that the user can identify it (e.g., "the vase on the table"). To be acceptable, these expressions should be "natural:" they should be at the same time sufficiently but not overly constrained, and they should not impose on the user a heavier cognitive load than necessary. For example, producing the expression "the vase that is not above the chair or sofa or under the table" would probably not be acceptable. Areces et al. (2008b) propose to use symbolic minimization of the model that represents the state of the world, in order to obtain a logical representation that describe each object uniquely. In our project we will implement this method and evaluate it within the dialogue system.

*(3) Management of the Interaction Context:* To manage the use of the interaction context we will use existing knowledge maintenance systems such as RACER[2] or Pellet[3], which support inference tasks such as definition, maintenance and querying of ontologies. These systems have been used as inference engines in numerous applications in the area and, in particular, in dialogue systems for text adventures (Benotti, 2009b). Once we have studied the behavior of these inference engines on the task, we will analyze its limitations and investigate the required extensions.

[2]http://www.racer-systems.com

[3]http://clarkparsia.com/pellet

*(4) Interpretation of User Responses:* The interpretation of user responses in the unidirectional system is relatively simple: it amounts to discretizing the continuous flow of user behavior in the 3D world into actions meaningful for the domain task. In a first stage, we will use the discretizer provided by GIVE. After evaluating it we can determine whether or not this module meets the requirements of our task and what are its limitations. In the bidirectional system, however, the interpretation of user responses is the task that will require more attention. To start with, the bidirectional system should be expanded with capabilities for processing statements coming from the user (namely, parsing, semantic construction, resolution of references, etc.). We will study, in particular, two types of user contributions: requests for clarification of the instruction given (what we call 'short-term repairs'), and for redefinition of goals (what we call 'long-term repairs'). We will implement short-term repairs using the approach described in (Purver, 2006). For long-term repairs we will use the guidelines of (Blaylock, 2005).

A sample interaction with the unidirectional system guiding the player in the identification of a particular blue button is as follows:

(1) System says: Push a blue button.
*The user focuses a blue button.*
System says: Not this one.
Look for another one.
*The user turns and focuses another blue button.*
System says: Yes this one!
*The user pushes the button.*

This interaction illustrates the tasks described above. To begin with, the verbalization of the instruction "Push a blue button" is making explicit one of the steps of the plan that needs to be performed in order to achieve the task goal. As we can see, the system implements in this case a referring strategy which does not uniquely identify the referent (the system generates "a blue button" when there is more than one blue button in the domain). But it is capable of producing further details about the referent if the user focus in the wrong object. Finally, this example makes evident that the interpretation of the user responses is crucial even in a linguistically unidirectional system. The user cannot make linguistic contributions but can change the context by performing physical acts, the correct interpretation of such acts is essential if the system is to react coherently.

### 2.3 Evaluation

To determine the quality of the obtained prototypes we propose to create a quality model following the ISO/IEC 9126 and 14528 standards for the evaluation of software products (ISO/IEC, 2001; ISO/IEC, 1999). These standards were successfully applied to the Machine Translation domain, resulting in the *FEMTI*[4], *Framework for the Evaluation of Machine Translation* (Estrella et al., 2005). FEMTI guides evaluators towards creating parameterized evaluation plans that include various aspects of the to-be-evaluated system and offer a relevant set of metrics. The identification of relevant metrics can be performed using various methods, e.g., based on previous experience (Hajdinjak and Mihelic, 2006; Litman and Pan, 2002), conducting surveys or requirement specifications (Lecoeuche et al., 1998), or collecting such data through Wizard of Oz experiments (Dahlbäck et al., 1998). After developing a quality model, several methodologies to assess various aspects of the system can be applied: automatic metrics, subjective metrics or metrics based on the task (to evaluate both the contribution of each component and the quality of the whole system).

The GIVE platform is used every year as a unified framework for evaluating generation systems. Systems have to generate natural language instructions and be able to participate in a real-time interaction situated in a 3D environment. The GIVE Challenge is one of the shared tasks endorsed by ACL's special interests groups in generation, dialogue and semantics. We plan to participate in the challenge, which will serve as an additional source of information about aspects of the system that need improvement. The evaluation metrics used in the Challenge (such as average reference identification time) are described in (Byron et al., 2009). In (Amoia et al., 2010) we extended such metrics in order to measure alingment between system and user. Once the prototype is evaluated and improved using the results of the challenge, we will investigate its use as a virtual language tutor as described in the next section.

[4] http://www.issco.unige.ch/femti/

### 2.4 An Application: A Virtual Tutor

The project outcome will be a system capable of giving natural language instructions situated in a virtual 3D environment. The technology and theoretical advances of the project could be used in various applications, but one of the most interesting characteristics we plan to investigate is that, a priori, by just changing the linguistic resources, the language of interaction with the system (input and output) can be changed as desired. After obtaining a first prototype of an instruction giving dialogue system, we will investigate its use for distance learning, adapting the system to operate as a foreign language tutor (Wik and Hjalmarsson, 2009).

A one-way system that generates instructions in English can be used to test the user understanding of a foreign language. The correct interpretation of the instructions can be evaluated from the proper execution of the instructions. The two-way system will allow the user to formulate clarifications (either in their native language or in the foreign language). The user may also redefine the objective to be achieved during the interaction, and thus select the type of vocabulary he wants to practice.

Virtual worlds (like Second Life) are being rapidly incorporated into education, both initial and superior (Doswell, 2005; Molka-Danielsen and Deutschmann, 2009). The use of a virtual tutor has certain advantages over a human tutor. Engwall (2004) mentioned the following. (1) Amount of practice: the chance to practice the new language is essential for learning, and a virtual tutor provides opportunities only limited by the technological resources. (2) Prestige: a student may feel embarrassed about making mistakes with a human tutor, and this might limit his willingness to speak in the foreign language. (3) Augmented Reality: a virtual tutor can provide additional material (e.g., examples in context, explanatory images, etc.) with less effort than a human tutor.

Such a virtual tutor can be used in distance learning. To develop distance learning systems, it is essential to model the user's learning progress. This requires a system aware of the evolution of the user, and that takes into account their achievements and their problems. The system must be able to interpret requirements, and generate appropriate responses, for non-experts uses whose knowledge evolves during the interaction. Moreover, the system must be able to properly represent both the information concerning the course material, and information about the evolution of the user. For example, the system must be able to diagnose what part of the course material should be reviewed from the wrong answers of the user. Finally, the system must be able to evaluate the user interaction in order to decide which learning objectives have been achieved. The theoretical and practical results of the project contribute to solving these difficult problems.

## 3 Impact of the Project

This project aims to achieve a balance between a system which is sufficiently generic to be applicable in different areas, and specific enough to benefit from the efficient use of existing techniques for knowledge management, planning and natural language processing. Designing and implementing such a system is a multidisciplinary effort leading to research in diverse scientific areas:

**Pragmatics** is an interdisciplinary field which integrates insights from linguistics (e.g., conversational implicatures (Grice, 1975)), sociology (e.g., conversational analysis (Schegloff, 1987)) and philosophy (e.g., theory of speech acts (Austin, 1962)). It aims to explore how the context (in which a conversation is situated) contributes to the meaning (of everything that is said during that conversation). The meaning conveyed during a conversation depends not only on linguistic information (entities in focus, grammatical and morphological rules, etc.) but also on extra-linguistic information (physical situation of conversation, previous experiences of speakers, etc.). As a result, the same sentence may mean different things in different contexts. The area of pragmatics studies the process by which a sentence is disambiguated using its context. A dialogue system needs to have pragmatic capabilities in order to interact in a natural way with its users. In particular, it must define what kind of contextual information should be represented; and what inference tasks on a sentence and context are necessary in order to interpret an utterance. In such a system it is important that sentences makes explicit the right amount of information: too much information will delay and bore the user, but if the information is not enough the user will not know how to perform the task and make mistakes.

One of the major contributions of the project in this area will be a virtual laboratory for pragmatic theories: a controlled environment for studying interaction set in a world where physical actions and language intermingle. The prototype will let us investigate the impact that different instruction giving policies (e.g., post order on the tree structure of the task) have on successful achievement of the goal. Similar studies have been done before (e.g., (Foster et al., 2009)) but they usually assume a predetermined task. Since our prototype allows for the specification of the virtual world, the available actions, and the goal, we will be able to determine when the impact associated to a particular policy is dependent on the task or not. We will also investigate short and long term repairs. Repairs are usually caused by conversational implicatures (Benotti, 2009a). Modeling these implicatures in a generic dialogue system is difficult because they are too open ended. However, since the present prototype provides a situated interaction, restricted to the virtual world, it will be possible to test the relationship between implicatures, the type of repairs they give rise to, and the inference tasks needed to predict them.

**Inference** can be understood as any operation that transforms implicit information in explicit information. This definition is general enough to cover tasks ranging from logical inference (i.e., deduction in a formal language) to inference tasks common in AI (e.g., planning and non-monotonic inference), as well as statistical operations (e.g. obtaining estimators on a data set). A dialogue system has to continually perform inference operations. E.g., inference is needed to interpret information received from the user, incorporate it to the system's data repository, and then decide what should be conveyed back to the user. The very problem of deciding what kind of logical representation and what type of inference to use in a given situation is complex (propositional logic vs. first-order logic, validity vs. model checking, logical inference vs. statistical inference). Independently of which type of inference is used, they are usually computationally expensive. The challenge here is to find the appropriate balance between the expressivity of the representation formalism and the cost of the required inference methods.

The main contribution of the project in this area is in the design, development and study of planning algorithms. A typical planning system takes three inputs –initial state, possible actions and expected goal– and returns a sequence of actions (a plan) that when sequentially applied to the initial state, ends in a state that satisfies the goal. Different methods to obtain a plan have been studied (forward chaining, backward chaining, coding in terms of propositional satisfiability, etc.), and they are currently implemented in systems that can solve many planning tasks efficiently. However, most of these systems make assumptions that simplify the problem (deterministic atomic time, complete information, absence of a background theory, etc.). And most of them return a single plan. We will investigate algorithms that eliminate some of these simplifications (in particular, we will study planning with incomplete information and based on a background theory). We will also provide extended planning services: alternative plans, minimal plans, conditional plans, incomplete plans, affordability of a given state, etc.

**Evaluation** of natural language generation systems is one of the most difficult tasks in the area of NLP. A given concept can be expressed in many different ways, all of them correct. Hence, it is not possible to determining the quality of a generated sentence simply by, for example, comparing the result with a gold standard. The problem of absence of gold standards is shared with another area of the NLP, namely Machine Translation, for which various evaluation methodologies, both direct and indirect, have been proposed. Direct methods applies a metric to the text generated by the system, while indirect methods evaluates the performance of the system through the use of the generated text to perform some task. But none of these methods is a standard and generally accepted methodology, which has been proven to be effective in all cases. Since what is being evaluated in this project is a system that interacts via the generation of natural language instructions, we can determine its performance through quantitative metrics (e.g., average task completion time), qualitative metrics (e.g., general user satisfaction) and metrics based on the context (e.g., how well the system addressed the user needs in particular situations). We will study the portability of evaluation techniques from the domain of machine translation and multimodal human-computer interaction to the evaluation of the system proposed in this project.

One of the main contributions of our project at this respect is the integration of assessment techniques from different areas into a methodology for evaluating dialog systems for virtual environments, aiming to estimate their usability and effectiveness. This methodology could be used both to determine whether a system is suitable for a task type and user, and to compare the performance of different systems of the same type. Another contribution will be the study and application of software evaluation standards to the developed systems, creating a standardized quality model and proposing a set of appropriate metrics to assess each of the aspects of the model. Finally, the annotated corpus of human-human interaction, together with the corpus of human-machine interaction collected during the project will be made public. Such corpora will serve, for example, to design more general platforms for evaluating dialog systems, going beyond the aspects evaluated by existing platforms like GIVE.

**Impact in the Argentinean Landscape:** Natural language processing, and in particular the field of dialogue systems is a rapidly growing area in developed countries. The automatic processing of natural language has become a strategic capability for companies and the wider community. However, this area is extremely underdeveloped in Argentina. This can be attributed to several factors. (a) The relative youth of the area of NLP, which implies a relative dearth of trained professionals throughout the world. (b) The underdevelopment of the area of research in Artificial Intelligence and Formal Linguistics in Argentina, for historical reasons and lack of industry demand. (c) Poor interaction between the few researchers in NLP that are in the region.

NLP is a strategic research area for Argentina which can achieve academic excellence and industry relevance. We believe in supporting the development of this area by promoting the following. (a) Training of human resources through doctoral programs and courses taught in Argentina by internationally renowned professionals. (b) Incorporation of trained human resources to contribute to the growth and diversification of the critical mass in the area. (c) Improving interaction between various groups and individual researchers in NLP, through the organization of workshops, courses, visits, co-tutoring, coordinated specialization programs, etc.

The particular topics investigated in the framework of this project are of relevance in the current Argentinean landscape for at least two reasons. On the one hand, the project integrates and develops various key aspects of the area of computational linguistics (syntax, semantics, pragmatics, representation, inference, evaluation); an area which, as we mentioned, is today almost nonexistent in Argentina. This project will be a step towards reversing this situation. On the other hand, the ultimate goal of the project is to investigate the use of the developed platform for distance education (specifically, as a tool for language learning). Distance education is a valuable resource to overcome the problem of centralization of educational resources in the country.

## 4 Introducing the Research Group

The PLN[5] research group, in which the describe scientific project will be carried out, was funded in 2005. Te group is developing an important role in human resource training, delivering courses to undergraduate and postgraduate student at the Universidad de Córdoba and other universities. It also works in the development of various research projects and integration with other groups in the region, both within Argentina and with neighboring countries (Chile, Brazil and Uruguay).

The current project pools together many of the key areas of expertise of the members of the group. To begin with, some members of the group specialize in computational logic, particularly in the theoretical and applied study of languages for knowledge representation (e.g., modal, hybrid and description logics). They have also developed automated theorem provers for these languages[6]. In relation with the study of knowledge representation, they have also investigated and developed algorithms for generating referring expressions (Areces et al., 2008b).

The second line of research of the PLN group that is relevant for this project is context-based evaluation. Members of the group have proposed an evaluation model for machine translation systems which relates the context of use to potentially important quality characteristics (Estrella et al., 2008; Estrella et al., 2009). This model is general enough to be applied to other systems that produce natural language like the ones proposed in this paper. Thanks to the background on machine translation systems the team has experience evaluating and comparing natural language output produced in different languages (Spanish and English in particular), which will be relevant for the development of the language tutor described in Section 2.4. Finally, the team has experience developing and evaluating multimodal corpora like those described in Section 2 (Estrella and Popescu-Belis, 2008).

The third line of research that is relevant for this project is pragmatics. In this area the team has implemented a conversational agent which is able to infer and negotiate conversational implicatures using inference tasks such as classical planning and planning under incomplete information (Benotti, 2009b). We have also investigated how to infer conversational implicatures triggered by comparative utterances (Benotti and Traum, 2009). Recently we have done corpus-based work, which shows what kinds of implicatures are inferred and negotiated by human dialogue participants during a task situated in a 3D virtual environment (Benotti, 2009a).

Other lines of research in the PLN group are not directly related to the project at this stage, but might become relevant in the future. They include grammar induction, text mining, statistical syntactic analysis and ontology population from raw text.

## 5 Ongoing and Future Collaborations

The members of the PLN in general and the authors of this paper in particular have several collaborations with national and international research groups in computational linguistics and related fields that are relevant for this project.

At the international level, we have ongoing collaboration with the TIM/ISSCO[7] *Multilingual Information Processing Department* at the University of Geneva, with the Idiap Research Institute[8] and

[5] http://www.cs.famaf.unc.edu.ar/~pln
[6] http://www.glyc.dc.uba.ar/intohylo/
[7] http://www.issco.unige.ch/en
[8] http://www.idiap.ch

with some members of the PAI[9], *Pervasive Artificial Intelligence* group of the University of Fribourg. These collaborations include the evaluation of NLP systems and the development of multilingual and multimodal human language technology systems.

Members of the group have a long standing collaboration with the TALARIS[10] group of the *Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA)*. The main research topic at TALARIS is computational linguistics with strong emphasis on semantics and inference. In the framework of this collaboration we are participating in the 2010 edition of the GIVE Challenge. In the process of designing the systems that will participate in the challenge we jointly investigated the use of different referring strategies in situated instruction giving (Amoia et al., 2010).

We have also collaborated with the Virtual Humans group of the Institute for Creative Technologies[11] from the University of Southern California. In particular we computationally modeled the inference of conversational implicatures triggered by comparative utterances (Benotti and Traum, 2009). The Institute for Creative Technologies offers Internship programs every year that we plan to use in order to strengthen our collaboration.

All these collaborations are directly related to the main theme of the project described in this article. The PLN group has also research collaborations with other international research teams in the framework of other scientific programs. For example, the PLN group has being part of a recently finished international project MICROBIO[12] on ontology population from raw text. The project was funded by the Stic-Amsud[13] program, a scientific-technological cooperation program integrated by France, Argentine, Brazil, Chile, Paraguay, Peru and Uruguay. The expertise obtained during this project might be useful in the future when trying to extend our GIVE ontologies to new domains. Similarly, the team maintain scientific relations with the University of Texas at Austin (mainly with Dr. J. Moore in projects related to the development of the ACL2[14] prover); and with the Research team Symbiose[15] of the Institut de Recherche en Informatique et Systémes Aléatoires (working on the use of linguistic techniques for the modelisation of genomic sequences).

At the national level, the group has intensively collaborated with GLyC[16], *Grupo de Lógica, Lenguaje y Computabilidad* on knowledge representation and inference (see, e.g. (Areces and Gorín, 2005; Areces et al., 2008a)). GLyC is part of the Computer Science Department of the Universidad de Buenos Aires. During 2010, teams PLN and GLyC will join forces and collaborate in the organization of ELiC[17], the *First School in Computational Linguistics* in Argentina, which will take place in July at the Universidad de Buenos Aires. ELiC 2010 will be co-located with the ECI[18], Escuela de Ciencias Informáticas which has a long standing reputation as a high-quality winter school in Computer Science in Argentina, and is being organized yearly since 1987. With ELiC we aim at creating, for the first time, a space to introduce the field of computational linguistics to graduate students in Argentina. Thanks to the support of the North American Chapter of the Association for Computational Linguistics (NAACL) and of the Universidad de Buenos Aires, ELiC is offering student travel grants and fee waivers to encourage participation.

The PLN group is also contacting other groups working in computational linguistics in Argentina like the research group in Artificial Intelligence from the Universidad Nacional del Comahue[19]. Taking advantage of previous co-participation in different project we plan to organize exchange programs in the framework of a research network.

Finally, the PLN group is planning to organize a workshop on Computational Linguistics as a satellite event of IBERAMIA 2010[20], the Ibero-American Conference on Artificial Intelligence, that will be organized by the Universidad del Sur, in the city of Bahía Blanca, Argentina.

[9] http://diuf.unifr.ch/pai/wiki
[10] http://talaris.loria.fr
[11] http://ict.usc.edu/projects/virtual_humans
[12] http://www.microbioamsud.net
[13] http://www.sticamsud.org
[14] http://www.cs.utexas.edu/users/moore/acl2
[15] http://www.irisa.fr/symbiose
[16] http://www.glyc.dc.uba.ar
[17] http://www.glyc.dc.uba.ar/elic2010
[18] http://www.dc.uba.ar/events/eci/2009/eci2009
[19] http://www.uncoma.edu.ar/
[20] http://cs.uns.edu.ar/iberamia2010

## References


M. Amoia, A. Denis, L. Benotti, and C. Gardent. 2010. Evaluating referring strategies in situated instruction giving. *Topics in Cognitive Science*. Submitted.

C. Areces and D. Gorín. 2005. Ordered resolution with selection for $H(@)$. In F. Baader and A. Voronkov, editors, *Proc. of LPAR 2004*, volume 3452 of *LNCS*, pages 125–141. Springer.

C. Areces, D. Figueira, S. Figueira, and S. Mera. 2008a. Expressive power and decidability for memory logics. In *Logic, Language, Information and Computation*, volume 5110 of *LNCSs*, pages 56–68. Springer.

C. Areces, A. Koller, and K. Striegnitz. 2008b. Referring expressions as formulas of description logic. In *Proc. of INLG-08*.

J. Austin. 1962. *How to do Things with Words*. Oxford University Press.

L. Benotti and D. Traum. 2009. A computational account of comparative implicatures for a spoken dialogue agent. In *Proc. of IWCS-8*.

L. Benotti. 2009a. Clarification potential of instructions. In *SIGDIAL-09*.

L. Benotti. 2009b. Frolog: An accommodating text-adventure game. In *Proc. of EACL-09*.

N. Blaylock. 2005. *Towards tractable agent-based dialogue*. Ph.D. thesis, University of Rochester, Department of Computer Science.

D. Byron, A. Koller, K. Striegnitz, J. Cassell, R. Dale, J. Moore, and J. Oberlander. 2009. Report on the 1st GIVE challenge. In *Proc. of ENLG*, pages 165–173.

N. Dahlbäck, A. Jönsson, and L. Ahrenberg. 1998. Wizard of Oz studies—why and how. In *Readings in intelligent user interfaces*, pages 610–619. Morgan Kaufmann Publishers Inc.

J. Doswell. 2005. It's virtually pedagogical: pedagogical agents in mixed reality learning environments. In *Proc. of SIGGRAPH-05*, page 25. ACM.

O. Engwall, P. Wik, J. Beskow, and G. Granström. 2004. Design strategies for a virtual language tutor. In S. Kim and D. Young, editors, *Proc. of ICSLP-04*, volume 3, pages 1693–1696.

P. Estrella and A. Popescu-Belis. 2008. Multi-eval: an evaluation framework for multimodal dialogue annotations. Poster at the Joint IM2 and ASSI.

P. Estrella, A. Popescu-Belis, and N. Underwood. 2005. Finding the system that suits you best: Towards the normalization of MT evaluation. In *Proc. of ASLIB-05*, pages 23–34.

P. Estrella, A. Popescu-Belis, and M. King. 2008. Improving contextual quality models for MT evaluation based on evaluators' feedback. In *Proc. of LREC-08*.

P. Estrella, A. Popescu-Belis, and M. King. 2009. The femti guidelines for contextual mt evaluation: principles and tools. In W. Daelemans and V. Hoste, editors, *Evaluation of Translation Technology*. Linguistica Antverpiensia.

M. Foster, M. Giuliani, A. Isard, C. Matheson, J. Oberlander, and A. Knoll. 2009. Evaluating description and reference strategies in a cooperative human-robot dialogue system. In *Proc. of IJCAI-09*.

P. Grice. 1975. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and Semantics: Vol. 3: Speech Acts*, pages 41–58. Academic Press.

M. Hajdinjak and F. Mihelic. 2006. The paradise evaluation framework: Issues and findings. *Computational Linguistics*, 32(2):263–272.

J. Hoffmann and B. Nebel. 2001. The FF planning system: Fast plan generation through heuristic search. *JAIR*, 14:253–302.

ISO/IEC. 1999. *14598-1:1999 (E) – Information Technology – Software Product Evaluation – Part 1: General Overview*.

ISO/IEC. 2001. *9126-1:2001 (E) – Software Engineering – Product Quality – Part 1:Quality Model*.

H. Kautz and B. Selman. 1999. Unifying SAT-based and graph-based planning. In *Proc of the IJCAI*, pages 318–325.

R. Lecoeuche, C. Mellish, and D. Robertson. 1998. A framework for requirements elicitation through mixed-initiative dialogue. In *Proc. ICRE-98*. IEEE.

D. Litman and S. Pan. 2002. Designing and evaluating an adaptive spoken dialogue system. *User Modeling and User-Adapted Interaction*, 12(2-3):111–137.

J. Molka-Danielsen and M. Deutschmann, editors. 2009. *Learning and Teaching in the Virtual World of Second Life*. Tapir Academic Press.

D. Nau, M. Ghallab, and P. Traverso. 2004. *Automated Planning: Theory & Practice*. Morgan Kaufmann Publishers Inc.

M. Purver. 2006. CLARIE: Handling clarification requests in a dialogue system. *Research on Language and Computation*, 4(2-3):259–288.

E. Schegloff. 1987. Some sources of misunderstanding in talk-in-interaction. *Linguistics*, 8:201–218.

P. Wik and A. Hjalmarsson. 2009. Embodied conversational agents in computer assisted language learning. *Speech Commun.*, 51(10):1024–1037.