Work in Intelligent Multimodal Dialogues in Spanish

Dr. Luis A. Pineda

1988 - 2008

Department of Computer Science
Institute for Applied Mathematics and Systems (IIMAS)
Universidad Nacional Autónoma de México

Dr. Pineda's work in multimodal interfaces and architectures for intelligent systems in AI started with the Graflog system developed during his PhD studies at the University of Edinburgh in 1988. This system had two independent representational structures for the representation of geometrical and conceptual knowledge, linked through a mapping interpretation function, and it was possible to express graphical configurations and interpretations through the multimodal interface, and ask questions about the facts stated and their consequences through a natural language and graphics facility supported with a pointing device. This work was originally presented in the paper "Understanding Drawings Through Natural Language" published in *Computer Graphics Forum* back in 1988, and was the subject of a series of papers and book chapters that he wrote in collaboration with his colleges at the University of Edinburgh over the next four years. The core of the interaction was a dialogue manager, which was defined in terms of a number of rules involving graphics and linguistic knowledge, and each rule had an associated reasoning or problem-solving action, that was performed by the system when the preconditions for such rule were satisfied.

Later on, at his incorporation to UNAM in 1998, Dr. Pineda founded the DIME group and started the DIME project with the purpose to construct intelligent multimodal dialogue systems in spoken Spanish, and he has continuously coordinated the group and developed this project to the present date. As an initial task the DIME group collected a multimodal corpus of task oriented conversations in spoken Spanish in the kitchen design domain, which was tagged at different linguistic levels, and has been used to study the phonetics, the intonation, and the syntax and pragmatics of conversational Spanish. Using this corpus as empirical data and the Head-Driven Phrase Structure Grammar (HPSG) formalism he developed a grammar of Spanish, which was focused on the structure of the periphrasis and the clitic system. This work was the subject of a series of paper by Dr. Pineda and one of his students and was presented in an extended version in the paper "The Spanish Pronominal Clitic System" in the Spanish journal *Procesamiento del Lenguaje Natural* in 2005. The DIME corpus was also used to study the structure of spoken conversations and Dr. Pineda, working with several students and collaborators over the years, developed the DIME-DAMSL tagging scheme for the analysis of speech acts in task oriented conversations. This work was also the subject of a series of conference papers that started in 2002 and culminated with the paper "The obligations and common ground structure of practical dialogues" published in *Inteligencia Artificial* in 2007, the Ibero-american journal of AI. The DIME Corpus was also the empirical base for a study on the relation between intonation and speech acts, that was developed in the context of his PhD student Sergio

Coria and, out of this work, they published the paper "An analysis of prosodic information for the recognition of dialogue acts in a multimodal corpus in Mexican Spanish" that appeared in *Computer Speech and Language* in 2008.

As a related effort and with the purpose to provide a speech facility to his conversational systems, Dr. Pineda started a line of work in speech recognition in Spanish. After some preliminary studies and prototype systems developed from 1998 to 2002, the DIME group started the design and collection of the DIMEx100 Corpus in 2004, which consists of a large set of spoken sentences with its phonetic transcription and pronunciation dictionaries. This effort involved a number of collaborators in Mexico, United States and Spain, and also a large number of students and technical staff that tagged the corpus manually over a period of four years. The project involved the adoption of a phonetic alphabet and the identification of the basic set of phonetic units for Mexican Spanish, and also the definition of the tagging conventions, to make this resource useful for the construction of speech recognition systems. The effort also involved the development of human and computational infrastructure for the construction of this kind of systems, which was based on the Sphinx software, developed at Carnegie Mellon University. As a result, the DIME group counts with a number of operational recognition systems and a flexible platform for the construction of Spanish speech recognition systems with potential applications to different domains, and it is perhaps the only group working in Mexico that has this infrastructure and capabilities. Also, the collection of the corpus was subject of a publication in 2004 that appeared in the Series *Lecture Notes on Artificial Intelligence*, and its transcription and validation is the subject of an extensive paper, which is still pending publication.

The work on the construction of multimodal architectures, speech act analysis and speech recognition, converged in the definition of a new dialogue manager. In this line and following the intuition that natural language is essentially a context depend process, Dr. Pineda developed the notion of dialogue models for the representation of the conversational context for specific application domains; dialogue models are schematic protocols defined at the intentional level that relate the intentions that conversational partners are expected to express in specific conversational situations with the actions that ought to be performed as a result of the interpretation of such intentions. For the representation of dialogue models he developed a formalism that he named "Functional Recursive Transition Networks". This consists of a notation and its program interpreter, and dialogue models are specified through direct acyclic graphs (DAGs) and represent conversational protocols and situations, with the parametric representation of the intentions and actions that define an application domain. A characteristic of this formalism is that intentions and actions are represented in a declarative and modality independent representational system. The dialogue model is the core of a multimodal architecture that has three levels of agents for interpretation and action: a first level for the recognition of modality specific information, like linguistic or visual, a second interpretation level, where intentions are interpreted in terms of information that flows bottom-up from the input agents but also top-down from the dialogue manager and dialogue models, and a third level constituted by the dialogue manager properly, which contains the modality independent specification of the interpretation context. In particular, the task of the intermediate level agents is to determine the most likely intention expressed in situation in terms of the modality specific information gathered through the modality specific recognition agents. Dr. Pineda

developed the program interpreter of this dialogue manager, and also a set of dialogue models to test the system. The notion of dialogue models and dialogue model interpretation was the subject of a lecture that Dr. Pineda gave in Soria, Spain, in 2004, where a first demonstration of the dialogue manager was also presented, and in 2005 he also developed a further version able to keep track of the conversation and use this information to interpret intentions in relation to previous interpretation acts too.

In parallel with this work and with the aim to integrate the multimodal conversation technology in situated agents, Dr. Pineda started the Golem project in 2002. In this project a mobile robot, that he named Golem, was used as a test bed for the multimodal interaction technology. The robot was first demonstrated in 2003 with a simple application based on finite state automaton, although this was very fragile and couldn't follow a simple conversation in a robust way. However, the development of the concept of dialogue models and its program interpreter in 2004 and 2005, and also a more reliable speech recognition infrastructure, permitted to develop a robust conversation al system that was demonstrated in 2006. In this application the Robot Golem was able to guide a poster session about the research projects developed at the Department of Computer Science at IIMAS, UNAM, through a multimodal conversation in spoken Spanish supported with other modalities, like text, images and video, and this demo has been demonstrated widely in Mexico. The notion of dialogue models and its associated dialogue manager was the subject of the paper "Specification and Interpretation of Multimodal Dialogue Models for Human-Robot Interaction", published last year as a book chapter. Also, the visual modality was added recently, following very closely the language interpretation module, and a demonstration in which the robot is able to talk about objects that are placed within its visual field is also available.