Journal of Virtual Worlds Research

jvwresearch.org ISSN: 1941-8477

MPEG-V & Other Virtual Worlds Standards

December 2011
Volume 4, Number 3

Volume 4, Number 3
MPEG-V and Other Virtual Worlds Standards
December 2011

This issue includes papers partially supported by the ITEA2 Metaverse1 Project (http://www.metaverse1.org).

Journal of
Virtual Worlds Research
jvwresearch.org ISSN: 1941-8477

Volume 4, Number 3
MPEG-V and Other Virtual Worlds Standards
December 2011

Associating Automatic Natural Language Processing to Serious Games and Virtual Worlds

Treveur Bretaudière, Samuel Cruz-Lara, and Lina María Rojas Barahona
LORIA (UMR 7503) – TALARIS team,
CNRS / INRIA / University of Lorraine, France

Abstract

We present our current research activities associating automatic natural language processing to serious games and virtual worlds. Several interesting scenarios have been developed: language learning, natural language generation, multilingual information, emotion detection, real-time translations, and non-intrusive access to linguistic information such as definitions or synonyms. Part of our work has contributed to the specification of the Multi Lingual Information Framework [ISO FDIS 24616], (MLIF, 2011). Standardization will grant stability, interoperability and sustainability of an important part of our research activities, in particular, in the framework of representing and managing multilingual textual information.

1. Towards a Standardized Linguistic Framework for Serious Games and Virtual Worlds

Linguistic information plays a crucial role in the management of information, as it undertakes most of the descriptive content associated with more visual information. Depending on the context, it may be considered as the primary content (i.e., text illustrated by pictures or videos), as documentary content for multimedia information, or as one among several possible information components in specific contexts such as interactive multimedia applications. Linguistic information can also appear in various formats: spoken data in an audio or video sequence, implicit data appearing on an image (i.e., caption, tags, and subtitles) or textual information that may be presented to the user graphically or via a text to speech processor.

The extremely fast evolution of the technological development in the sector of Communication and Information Technologies, and in particular, in the field of natural language processing, makes particularly acute the question of standardization. The issues related to this standardization are of an industrial, economic and cultural nature. Nowadays, an increasing number of standards are frequently being used within most scientific and technical domains. In previous work (Cruz-Lara S., Bellalem N., Bellalem L., and Osswald T., 2009; Cruz-Lara et al., 2011) we have explained how a standardized framework for textual multilingual support (MLIF, 2011) associated with an immersive 3D environment, may considerably change the way people usually deal with multilingual information and with language learning.

Our current research activities are mainly related to the following projects: ITEA2 METAVERSE1 (http://www.metaverse1.org), INTERREG ALLEGRO (http://www.allegro-project.eu/index.php?id=13766), and EUROSTAR EMO-SPEECH (http://www.eurostars-eureka.eu/search.do?method=display&id=10). What we would like to do in this paper, based on the results we have obtained in these projects, is to setup foundations to a general linguistic framework for video games or virtual worlds.

2. The EUROSTAR EMO-SPEECH Project

1.1. Supporting conversations for serious games

Serious games engage players towards the acquisition of new skills rather than merely entertain them. Typically, they are designed for education, training or general counseling and assistance. Virtual Characters (VCs) or Embodied Conversational Agents (ECAs) can enrich users` engagement through interactive conversations (Cassell J. et al., 1999). For instance, VCs can help players to achieve some goals inside the game, such as, orienting players to find a certain place in a 3D environment, or providing relevant information at each level of the game. Cognitive models have been proposed for representing conversational agent beliefs, desires and intentions. The levels of thrust, credibility and cooperativeness have been subject to a study in the last years, in which VCs can be collaborative or not (Roque, A., and Traum D.R., 2007). Our aim in the EUROSTAR EMO-SPEECH project is to provide virtual characters in serious games with conversation capabilities in collaborative environments.

3. Dialog Modelling

Multimodal dialogue systems are complex distributed and asynchronous architectures that gather specialized components. Broadly, these components solve the tasks of modal-based recognition and synthesis, understanding, dialogue management, generation, fission and fusion; and they can be either symbolic or stochastic oriented. The lack of domain-specific and linguistic resources is the major difficulty when incorporating dialogue in different domains and languages.

In order to support conversations in serious games, we implemented a Dialogue Server that follows the Information State approach. The notion of information-state (IS) is somehow known in the literature as "conversational score", "discourse context" or "mental state" (Larsson S., and Traum D.R., 2000). This approach maintains a blackboard structure with the relevant information. The information-state, (e.g. input, system beliefs, shared beliefs or common ground, etc.) identifies the manner in which this information is modified during interactions in the dialogue (i.e. update rules), and selects the system response taking into account a strategy (i.e. selection rules). We integrated and adjusted the state-of-the-art dialogue manager Midiki (Burke C. et al., 2003) within a distributed architecture: the Open Agent Architecture (OAA) (Cheyer A., and Martin D., 2001), for connecting heterogeneous linguistic components. We modeled goal-oriented conversations, in which, virtual characters know the type of knowledge the player must acquire to successfully complete a task in the game and the optional information the player should know to increase her score.

The dialogue itself can be seen as a game in which each speaker makes a move at each turn, namely a dialogue move (Larsson S., and Traum D.R., 2000). Dialogue moves are represented as predicates in a rule-base dialogue manager, in which the moves are resolved by unification. There is a plan (or strategy) for each dialogue that contains a sequence of system moves. The dialogue manager, checks the plan to find the first thing to do, the first system move is then translated to natural language by an utterance-selection generate component. It waits for player inputs. Once the player says something, the information state is updated and the interpreter finds the most plausible user move. Afterwards, the dialogue manager executes a series of selection rules against the information-state and the plan in order select the following system move and so forth.

In our first case study, we integrated this dialogue framework with the serious game "Mission Plastechnologie"[1] developed by Artefacto[2] (Figure 1). In this game, the player has to manufacture a joystick in a Pastech Factory. The employees of the factory are very collaborative and help the player to manufacture it. They explain the production and recycling processes, the different jobs and the security measures. Several characters are engaged in twelve different dialogues during the game[3]. Particularly, the roles of the player and the system are interchanged during the game depending on the game level. The player can be represented by three possible characters: Julie, Lucas, and Ben (Figure 1, right-hand side).

1 Mission Plastechnologie. http://www.mission-plastechnologie.com/

2 Artefacto. http://www.artefacto.fr/en/news/?id=last

3 The game and its dialogues are in French.

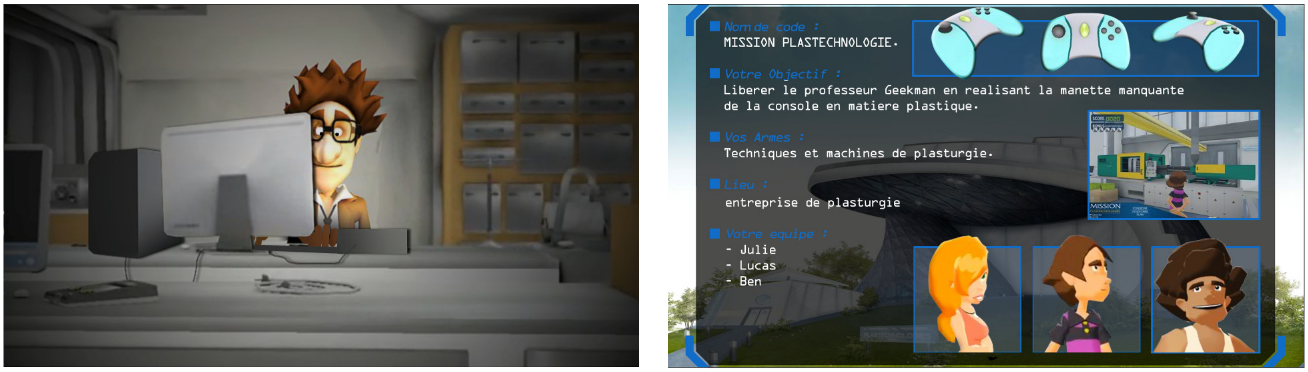

Figure 1: Scenes of the game "Mission Plastechnologie"

1.2. Data Collection and Analysis

We collected French utterances for all the dialogues in the game. We implemented two configurations of "wizard-of-oz", in which a human simulates some linguistic components within a distributed architecture. The first configuration, the free-wizard, supports a simple conversation between the player and the wizard. Here the wizard acts as the interpreter, dialogue manager and generate components. There is a virtual dialogue manager that solely retrieves specific information of the dialogue in a database, for instance the goals by dialogue. However, the dialogue strategy is completely free and the wizard has the responsibility of leading the conversation. The second configuration, the semiautomatic-wizard, integrates the wizard contributions within the dialogue framework explained before (Figure 2). It is worth noting that the semiautomatic configuration is a dialogue system, in which the dialogue manager controls the strategy and the wizard acts as the interpreter and generate components. In both cases the wizard-of-oz was implemented as a specialized linguistic component within a multi-agent architecture[4].

We organized a set of experiments with the semiautomatic wizard, since our immediate goal is to integrate system-driven dialogues in the game. We invited around 40 people to play the game and we collected 591 dialogues, 4874 utterances and 77854 words.

We are currently analyzing the collected corpus, according to the player moves in order to implement a machine-learning interpreter that returns the most plausible moves given a player sentence.

4 The Game Agent and ASR components in the architecture have been implemented by the PAROLE Team from LORIA.

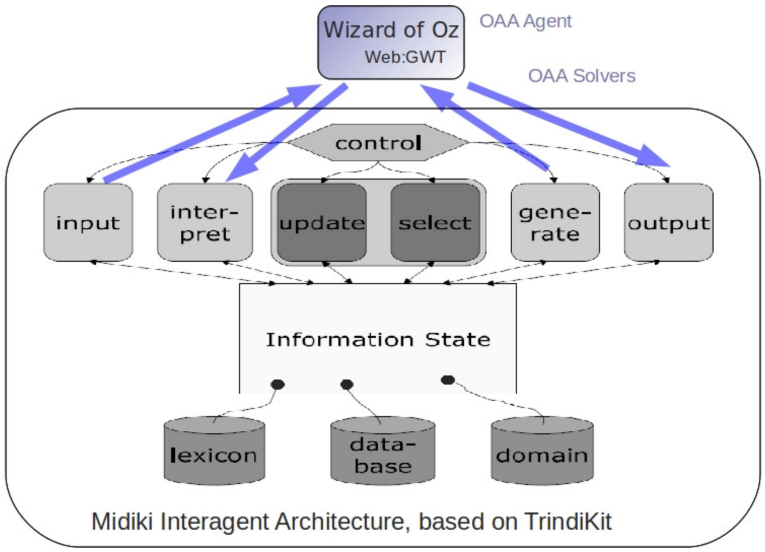

Figure 2. The semiautomatic wizard configuration.

The wizard interprets player utterances and edits system responses.

4. The INTERREG ALLEGRO Project

The INTERREG ALLEGRO project aims at making language learning more challenging and entertaining. It attempts to differentiate itself from the regular serious gaming market by relaying extensively on Natural Language Processing (NLP) systems. The introduction of these technologies allows for very sophisticated language learning exercises, contrasting with previously developed products, based on key words, template based generation, or a choice of fixed answers.

Thanks to language generation, the player can be confronted with many different ways to perform the very same exercise, increasing the player's interest, lowering her/his ability to remember previous exercise answers, and adapting the difficulty to the player's level; in short: it raises the challenge to another level.

This versatility enabled the use of dynamic environments, and this is why virtual universes were naturally chosen as a base for the game environment, allowing for a broader audience, easy maintenance, and cut on development costs in comparison with the costs of a dedicated 3D application.

A particular architecture has been designed to make the virtual environment a dynamic extension of the NLP (Figure 3). The software at the Allegro server side is a Web Ontology Language (OWL) based Java front-end to various NLP systems. The Allegro server sided architecture has been previously described in detail (Amoia M., 2011; Amoia M., Gardent C., and Perez-Beltrachini L., 2011), this is why we focus on the virtual environment side of the system. The general architecture of the project is composed of three main layers:

- NLP systems.
- Cross-technologic transaction systems.
- A virtual world related software frame for the game.

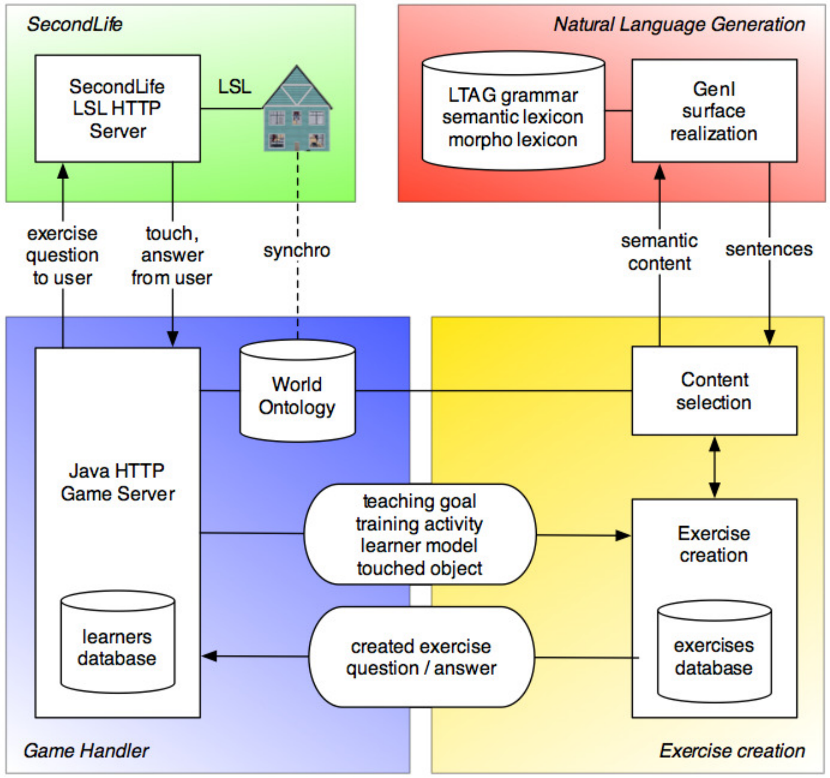

Figure 3: The general architecture of the INTERREG ALLEGRO project5

The implementation of the virtual environment tools and framework occurred in three different steps. We first had to integrate, in a transparent fashion, the NLP framework into the Second Life and OpenSIM virtual worlds, design the software architecture at the virtual world level, and finally determine a set of tools to easily extend the linguistic and knowledge resources.

We cannot reproduce the tremendous complexity of NLP systems in a virtual universe, therefore we had to establish an HTTP bridge between the two servers, the LSL one in the Multi User Virtual Environment (MUVE) and the Java one at the Allegro server side, in charge of the user credentials and the NLP various software.

To make the virtual environment a reliable and reactive dynamic extension of the NLP, a set of operands has been determined to exchange the game data between the MUVE and our servers. The idea was to provide the users with NLP based "serious gaming" exercises, confront and correct their answers whenever necessary, gather their results, export the best players scores, without delays between our servers and the MUVE to preserve the user's experience.

When a user plays, the Java system should at all times be aware of a user's credentials, determine the score of a user for a game, track the previous results of a player, gather questions at the NLP side, broadcast it to the MUVE, that in turn will broadcast the players results back, for the NLP to analyze its content and determine if the answer is valid, or provide a correction otherwise, and whenever necessary.

[5] Diagram from Alexandre Denis, TALARIS team, LORIA.

When creating a serious game, the exercise logic isn't the only aspect of the playability. The players should be provided with an environment featuring a coherent and not too complex user interface that would steep the systems' learning curve, allowing for gamers to totally focus on learning instead of losing time in understanding the software.

We did reduce the interaction to the minimum: users can click on objects to perform exercises that are generated in real-time according to the user's level, and current exercise mode. All the information is displayed on screen thanks to a heads up display featuring UTF-8 support for multilingual compatibility.

Considering the amount of objects in a sole environment, such as houses (around 200), we had to take the modification process of all objects into consideration. This modification process must be as easy as possible and straightforward.

Moderators can easily define the ontological properties of objects, thanks to simple dialogues and requesters that broadcast any modification or addition in the instance of an object to the Allegro servers; the ontology of an object ultimately allows for the NLP software to generate very precise exercises related to it.

The in-world architecture is articulated in a centralized way. The "environments" encompass a series of related objects, closely related to similar real-life environments, such as a house. Each environment is split into areas, like a kitchen or a room for ontological reasons. Finally all the objects are defined under a particular group, that allows differentiating not only environmental and area results, but also provide duplicates of a very same "environment" set in a different graphic landscape, to check if it has an impact over the players results.

We faced several challenges upon setting up our game infrastructure into the MUVE:

- Security aspects. Such environments are opened to the public, therefore should bear protections against repetition attacks, data insertion and data modification. We also established a unique transaction and user data broadcasting channels as a means to preserve the user privacy.
- MUVE responsiveness should be maintained at all times. This is why the software is highly optimized to generate as little lag as possible. It also avoids all sorts of in-world data storage to cope with the tight software memory window.
- We should also favor user bound transactions through maintaining the highest usability and availability of the system to preserve the gamers' user experience.
- Error recovery handling is equally important, at many levels from user session management, HTTP management, server errors or grid instability. Everything is done to handle minor issues transparently, so that players never get confronted to system downtimes.
- In-world communications are established in a secure way, through several layers of communication and software handshakes. In certain cases the communications are distributed.
- Simple configuration and auto configuration procedures to deal with the creation, duplication and replication of environments and objects.
- Updatability. Updating the software at high speed within thousands of different 3D objects needs serious communication and security protocols.

We are now in the implementation phase of the project. A major part of these issues is currently being addressed.

5. The ITEA2 METAVERSE1 Project

The ITEA2 Metaverse1 project provides a standardized global framework enabling the interoperability between virtual worlds (as for example Second Life™, World of Warcraft™, IMVU™, Active Worlds™, Google Earth™ and many others) and the real world (sensors, actuators, vision and rendering, social and welfare systems, banking, insurance, travel, real estate and many others).

In order to bridge the differences in existing and emerging metaverses, MPEG-V Media Context and Control (ISO/IEC 23005) will provide a lower entry level to (multiple) virtual worlds both for the provider of goods and services as well as the user (Timmerer, C., Gelissen, J., Waltl, M., & Hellwagner, H., 2009).

In the ITEA2 METAVERSE1 project we focused on developing linguistic tools that, among others, may help people to overcome certain language barriers.

1.3. Emotion Detection

The emotional aspect of communication is often lacking in the textual conversations in virtual worlds. Whereas textual chat is apparently the most common way to communicate in virtual worlds, the studies about emotion detection mostly concentrate on facial, audio and gestural emotions (Kao et al., 2009).

In the frame of the Metaverse1 project, the TALARIS team is in charge of developing a prototype of a new emotion detector based on textual inputs. Various ways have already been explored (Neviarouskaya, A., Prendinger, H., and Ishizuka, M., 2009), but we wanted to explore different pragmatic and empiric methods, in order to be able to categorize different kinds of sentences (Lu, C.-Y., Hong, J.-S., and Cruz-Lara, S., 2006).

The aim is to develop a remote web service for the text analysis and emotion detection, which is independent from any virtual world. The rendering part, of course, is directly managed by the virtual world:

- In Second Life, a scripted object sends the request to the web service and calls the appropriate Animation Override (AO) on the avatar, with respect to the returned emotion;
- In Solipsis[6], Artefacto integrates the request directly in the virtual world server, which is then in charge of rendering the emotion on the avatar.

To render the emotions in Second Life, we have created a wristband which changes colors depending on the emotion. The wristband embeds the script that calls the web service and runs the animation overrides, as is visible on figure 4.

[6] Solipsis. http://www.solipsis.org

Figure 4: **Emotion detection with a wristband and rendering in Second Life**

1.4. The Multilingual-Assisted Chat Interface

By developing the multilingual-assisted chat interface, we intended to provide new features to the chat interfaces, especially in the virtual worlds:

- Coloring words depending on their grammatical category (Figure 5);
- Providing information about words (synonyms, definitions, translations);
- Being able to have several multilingual conversations at the same time.

The purpose of the multilingual-assisted chat interface is twofold.

One target is language learning for people who want to improve their foreign language skills by exercising in immersive situations. We consider that these people already have some knowledge of the language they want to learn. But as they cannot be considered as having a fluent level yet, they may still need assistance in understanding some words or sentences.

The other targeted population of the multilingual-assisted chat interface is people who want to communicate with someone else, but cannot speak any language in common and do not necessarily want to learn the other avatar's language.

Figure 5: Coloring words in the multilingual-assisted chat interface (Second Life)

In both cases, the interface is designed in order to help people talk to one another, even if they do not share the same language.

We can distinguish two main functionalities in the multilingual-assisted chat interface and thus, we can imagine two different scenarios:

- David is a French student, having a medium English level. But he is very interested in this language and in meeting other people. In a virtual world, David is using the multilingual-assisted chat interface, which is going to help him very much. David is sightseeing when an avatar starts speaking English to him. Thanks to the special chat interface, David enables the coloration of verbs, so that all the verbs are displayed in a red font color. When he reads a word he does not know yet, he has the choice of directly clicking on it to retrieve some synonyms, definitions and translations of that word from remote corpora. His requests are stored in a database, so that he can review all of his requests at any time and learn from them.
- Michel, a French student, cannot speak any foreign language but is interested in Japan. While visiting a virtual town where many Japanese are hanging out, he meets Seta, who cannot speak any language other than Japanese. Fortunately, Michel uses the multilingual-assisted chat interface, which is going to translate all the sentences he receives into French. When he speaks, his sentences are translated to the most recent language in which Michel got a message. Doing so, Seta can understand Michel’s sentences even if he is not using the multilingual-assisted chat interface. Moreover, if Michel wants to see some Japanese writing, he is still able to click on a translated sentence and see the original sentence, as Seta wrote it.

The multilingual-assisted chat interface has been developed using web applications, connected to external web-services and linguistic corpora (WordNet, ConceptNet, Google Translate, etc.) interfaced with virtual worlds (in our case Solipsis and Second Life).

Another important point is the way in which the exchanged data has been standardized, in order to be easily adapted to any platform, especially any virtual world. In fact, most of the data exchanges rely on the Multi Lingual Information Framework – MLIF [ISO FDIS 24616] (MLIF, 2011).

The translation quality depends on the length of the texts. Most of the time, the shorter the sentence, the more reliable the translation. It should also be noted that the dialogues have to be written in an academic spelling.

6. Conclusion

We presented our current research activities associating automatic natural language processing to serious games and virtual worlds. This association has allowed interacting naturally with serious games (EUROSTAR EMO-SPEECH), to setup new paradigms for foreign language learning in virtual worlds and serious games (INTERREG ALLEGRO), and to develop multilingual non-intrusive interactive frameworks and emotion detection from textual information for virtual worlds (ITEA2 METAVERSE1).

Not only do virtual worlds and serious games provide a new access to information and give new dimensions to interaction, but they also allow the creation of innovative situational contexts and scenarios. Associating automatic natural language processing to serious games and virtual worlds enhance their potential use in the domain of education and training.

7. Discussion and Future Work (EMO-SPEECH).

We designed wizard-of-oz experiments for collecting dialogues in French and we gathered an important number of dialogues. Now, we are planning to incorporate a machine-learning interpreter in the current dialogue server in order to support completely automatic dialogues. Moreover, we are refining the dialogue manager strategy and the generator component based on the corpus. Experiments with the free-wizard will permit us to study mixed-initiative dialogues for supporting more natural interactions in the future.

8. Discussion and Future Work (ALLEGRO).

The integration of NLP into virtual universes is nearly done, and the results are promising, especially if we consider that a lot of information systems and frameworks can be stacked over the current framework. The current architecture already enables us to extract usage information which will enable measuring the software adaptability for its planned purpose. Ultimately, the amount of sustained gamers will determine the success of the whole.

9. Discussion and Future Work (METAVERSE1).

The multilingual needs are increasing every day and virtual worlds are a very good example of fields of development where applications that support multilingualism are becoming absolutely necessary. In order to enhance the interoperability between virtual worlds, applications and corpora, it is obvious that standards should come into force. This is the main goal of MLIF and, on a more global perspective, of the Metaverse1 project.

References

Amoia, M. (2011). I-FLEG: A 3D-Game for Learning French. *Proceedings of the 9th International Conference on Education and Information Systems, Technologies and Applications* (EISTA 2011). Orlando, Florida, USA

Amoia, M., Gardent, C., and Perez-Beltrachini, L. (2011). A serious game for second language acquisition. *Proceedings of the Third International Conference on Computer Aided Education* (CSEDU 2011), Noordwijkerout, The Netherlands.

Burke, C., Doran, C., Gertner, A., Gregorowicz, A., Harper, L., and Korb, J., et al. (2003). Dialogue complexity with portability?: research directions for the information state approach. *Research Directions in Dialogue Processing Workshop at the 2003 HLT-NAACL/NSF Human Language Technology Conference*.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., Yan, H. (1999). Embodiment in Conversational Interfaces: Rea. *ACM CHI 99 Conference Proceedings*, Pittsburgh, PA

Cheyer, A., and Martin, D. (2001). The Open Agent Architecture. *Autonomous Agents and Multi-Agent Systems. Volume 4, Numbers 1-2, 143-148*. Kluwer Academic Publishers.

Cruz-Lara, S., Bellalem, N., Bellalem, L., and Osswald, T. (2009). Immersive 3D Environments and Multilinguality: Some Non-Intrusive and Dynamic e-learning-oriented Scenarios based on Textual Information. *The Journal of Virtual Worlds Research 2, 3*

Cruz-Lara, S., Osswald, T., Guinaud, J., Bellalem, N., and Bellalem, L. (2011). A Chat Interface Using Standards for Communication and E-learning in Virtual Worlds. *Lecture Notes in Business Information Processing 73, Part 6, pp. 541-554*

Kao, E. C.-C., Liu, C.-C., Yang, T.-H., Hsieh, C.-T., and Soo, V.-W. (2009). Towards Text-based Emotion Detection. International Conference on Information Management and Engineering (pp. 70-74).

Lu, C.-Y., Hong, J.-S., and Cruz-Lara, S. (2006). Emotion Detection in Textual Information by Semantic Role Labeling and Web Mining Techniques. Third Taiwanese-French Conference on Information Technology.

Larsson, S., and Traum, D.R. (2000). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *The Journal of Natural Language Engineering, Volume 6 Issue 3-4*. Cambridge University Press New York, NY, USA

MLIF (2011). The Multi Lingual Information Framework (MLIF) - ISO FDIS 24616. *The International Organization for Standards*.

Neviarouskaya, A., Prendinger, H., and Ishizuka, M. (2009). EmoHeart: Conveying Emotions in Second Life Based on Affect Sensing from Text. Advances in Human-Computer Interaction, volume 2010, September 2009

Roque, A., and Traum D.R. (2007). A model of compliance and emotion for potentially adversarial dialogue agents. *In Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*.

Timmerer, C., Gelissen, J., Waltl, M., & Hellwagner, H. (2009). Interfacing with Virtual Worlds. *Network and Electronic Media Summit*