

# Verbal and Written Interaction in Virtual Worlds

## Some application examples

Pierre Nugues

Institut des Sciences de la Matière et du Rayonnement

6, boulevard du Maréchal Juin

F-14050 Caen, France

[pnugues@greyc.ismra.fr](mailto:pnugues@greyc.ismra.fr)

<http://www.greyc.ismra.fr/~pierre>

### ABSTRACT

This text first summarizes what can be the respective advantages of language interaction in a virtual worlds and 3D images in language interactions and dialogue.

It then describes three examples of verbal and written interaction systems in virtual reality, starting with Ulysse, a conversational agent that can help a user navigate in virtual worlds. Ulysse has been designed to be embedded in the representation of a participant of a virtual conference. Ulysse responds positively to motion orders and navigate the user's viewpoint on his/her behalf in the virtual world. On tests we carried out, we discovered that users, novices as well as experienced ones have difficulties moving in a 3D environment. Agents such as Ulysse enable a user to carry out navigation motions that would have been impossible with classical devices.

The second example is a prototype to recreate car accidents in a virtual world from written accident reports. Reports have been supplied by an insurance company and describe most often collisions between two vehicles. We could animate scenes, coordinate entities in the virtual world and thus replay some of the accidents from their natural language descriptions. Such animation would have been difficult using classical interaction devices. The text describes techniques we have used to implement our prototype and the results we have obtained so far. It explains how an information extraction system can benefit from such a tool.

Finally, the text describes a virtual workbench to study motion verbs. From the whole Ulysse system, we have stripped off a skeleton architecture that we have ported to VRML, Java, and Prolog. This skeleton allows the design of language applications in virtual worlds.

**Keywords:** Virtual reality, Conversational agents, Spoken navigation, Scene generation from texts.

### 1 COMPUTER INTERFACES AND VIRTUAL REALITY

Computer interfaces have now widely stabilized into a set of paradigms where variations are merely cosmetic. Following the Macintosh's finder, Windows' desktop or X-Window's avatars have all converged into visualization and interaction means that include windows, icons, menus, and pointers – the so-called WIMP model.

This model may have reached a plateau and be unable to scale up to new computing entities such as the Internet or to be adapted to new tasks such as cooperative work or simulation.

From ideas to recreate desktop tools on the screen of a computer under the form of symbols – icons –, some researchers thought that virtual reality was a better paradigm to design metaphors. This pushes desktop symbolization closer to reality and presents a way to escape some trends of the GUI routine. In applications such as computer tools for collaborative work others researchers saw virtual reality as means to situate users, to make them aware of the context: their coworkers and other working teams notably, and to enable easier communications [[1]].

In many respects, virtual reality is appealing because it brings more realistic images and more interaction to the computer desktop. Although cognitive values of visualization and visual think have been extensively described and researched [[2], [3], [4]], power of images and interaction is probably best captured by the Chinese proverb:

*I hear and I forget, I see and I remember, I do and I understand.*

And virtual reality addresses the two last points better than any other interface. From this viewpoint, virtual reality would appear as the extreme trend of the interactive desktop metaphor and an ultimate interface.

## 2 COMPUTER INTERFACES AND LANGUAGE

In virtual reality environments however, navigation is often difficult and interaction is can be oppressive. It extends in that way drawbacks of existing interfaces as well as their advantages. Virtual reality requires extensive and sometimes tricky gestures. Opening a folder in the recreated office of a virtual world would probably require more movements than with the Macintosh Finder.

Information access, desktop control which are often tied to ease of navigation are key points of a good interface design. It's not sure that widely available virtual reality interfaces address this problem. In experiments we conducted earlier, we found that computer novices as well as experienced computer users – but unfamiliar to virtual reality – had much navigational difficulties in virtual worlds [[5]]. In addition, virtual reality embodies the principle of direct interaction that requires from objects to be visible, close, and in a relatively upright position. When not visible, objects are sometimes difficult to find and then to approach which adds a supplement of navigation chore.

A language interface would enable easier designation and navigation and hence help users complete their tasks faster. If this last statement has no definitive proof, a hint can be given by the analogous example of the Web development that shows that most popular portals are natural or

constraint language interfaces (e.g. Altavista, Voilà, Excite, Lycos, Yahoo). Thanks to their indexing robots they prevent a user from clicking zillions of pages before finding relevant information. Although, the extent of language processing behind these sites might be discussed, it shows a clear user preference to designate things using Dutch (or French) rather than to navigate links.

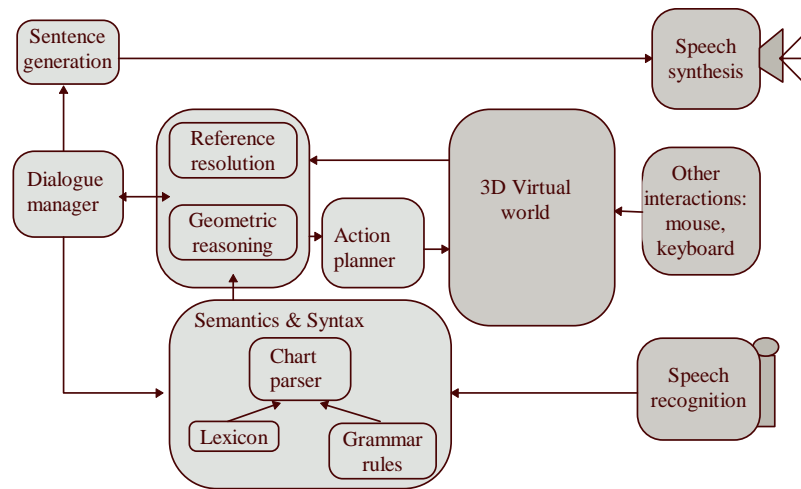
Gentner and Nielson [[6]] in a prospective article underlined limits of the WIMP interfaces. They described the possible role of language in future interfaces. They noted that language lets us refer to objects that are not immediately visible, encapsulate complex groups of objects, and support reasoning. They predicted a slow pervasion of natural language techniques in interfaces that would allow a negotiation between the user and its interface thanks to limited linguistic capabilities.

In our recent projects, we have implemented natural language agents in an attempt to bring to virtual reality systems some advantages of linguistics capabilities. That's what we describe now.

## 2. ULYSSE

Ulysse was our first implementation of a linguistic device in a virtual world. From user studies that we undertook, we discovered that many users were not able to move properly in a virtual world. We designed and implemented a conversational agent to help their navigation [[5], [7], [8], [9], [10]].

Ulysse consists in a chart parser and a semantic analyzer to build a dependency representation of the word stream [[11]] and a case form. It also features a reference resolver to associate noun phrases to entities of the virtual world and a geometric reasoner to cope with prepositions, groups, spatial descriptions, and to enable a limited understanding of the structure of the virtual world (Figure 1).



**Figure 1 Ulysse's architecture.**

Ulysse is embedded in the representation of the user in the world. Upon navigation commands from the user, Ulysse analyzes the word stream and navigates the user's viewpoint on his/her behalf in the virtual world. Ulysse's action engine is a planner that

uses an algorithm derived from STRIPS [[12]] (Figure 2). Ulysse can be used with a keyboard interface or a speech recognition system such as IBM's VoiceType or ViaVoice.

### 3. ULYSSE TO NAVIGATE INTO A BRAIN

We implemented the first version of Ulysse using the DIVE virtual environment [[13]]. We ported it to a PC to navigate into a reconstructed brain [[14]]. We kept the Ulysse overall architecture, but we had to modify the image display. Our idea was to combine realistic images and dialogues to explore brain regions and their functions [[15]].

The Cyceron research Center in Caen provided us with magnetic resonance images of the brain. We extracted 3D surfaces from them using thresholding operations. We chose arbitrary points on the surface that we associated to an arbitrary color. The colors were then diffused on the surface.

We investigated possible interactions and we designed scenarios to manipulate and navigate into this virtual brain. The scenarios have been designed in cooperation with the art group *Das synthetische Mischgewebe*. These scenarios have been limited to consider main regions of the brain such as the hemispheres, frontal lobes, etc. We restrained navigation to carry out linear motions and rotations relative to a designated object.

The action manager enables animation such as the sequence in Figure 3 that correspond to the utterance:

*Je voudrais voir le tronc.*

*I would like to see the medulla*

Although this prototype has only been used for art performances, it could find other applications. They include education and the interactive discovery of the brain anatomy and functions.

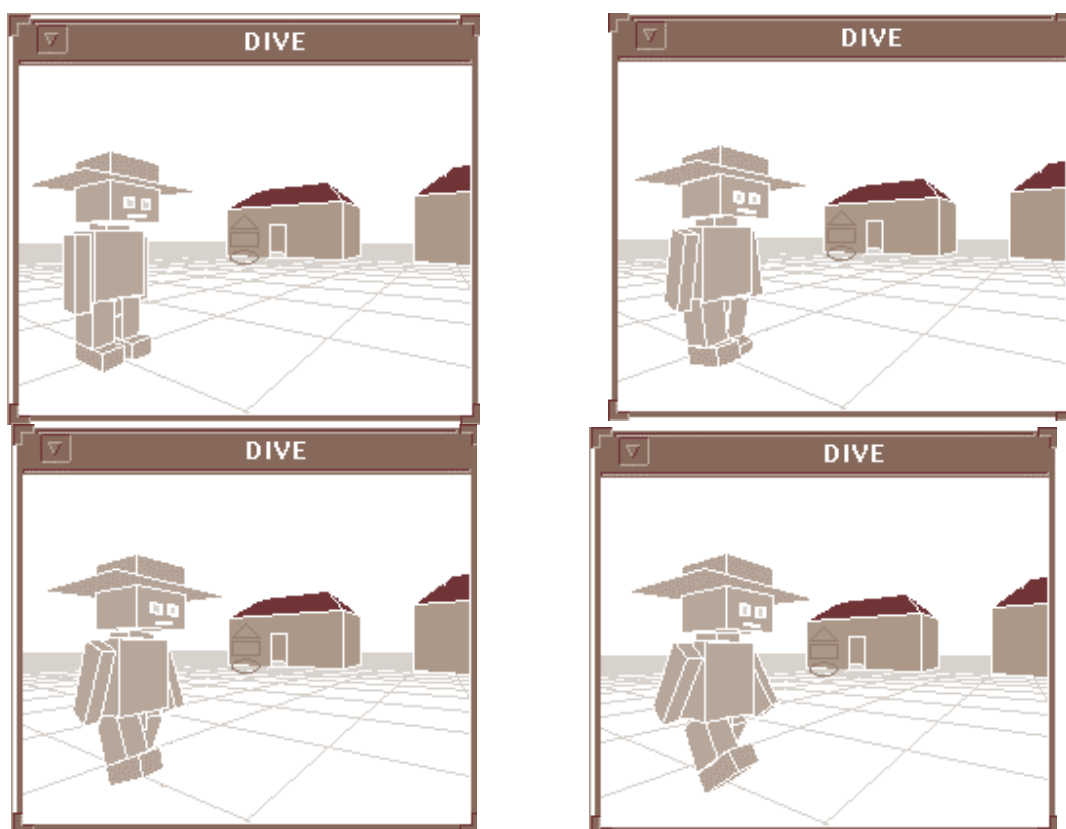
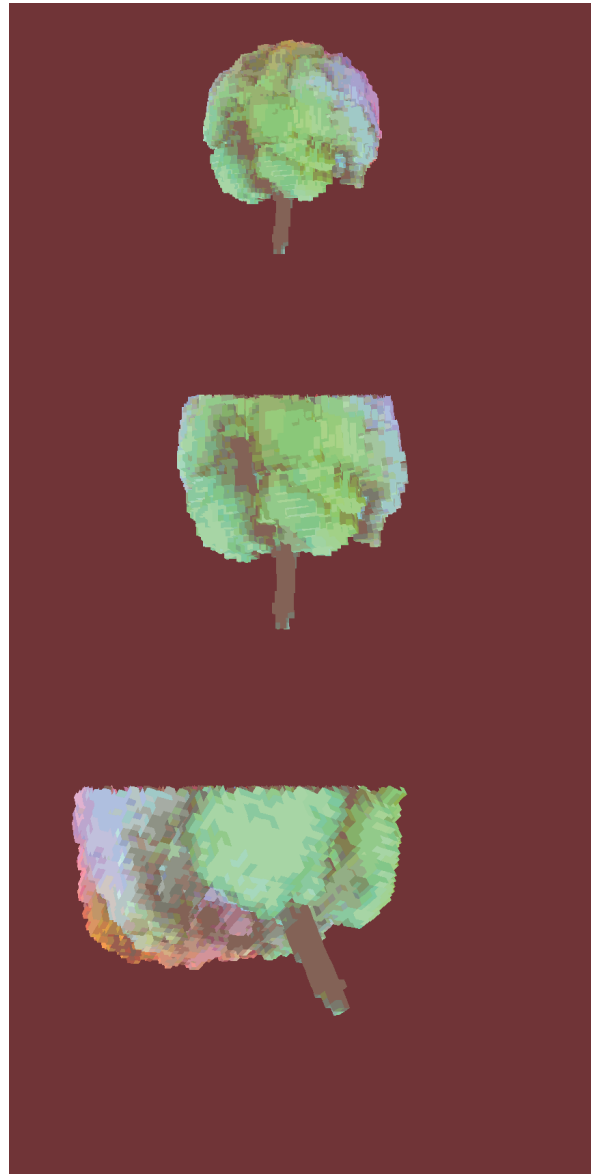
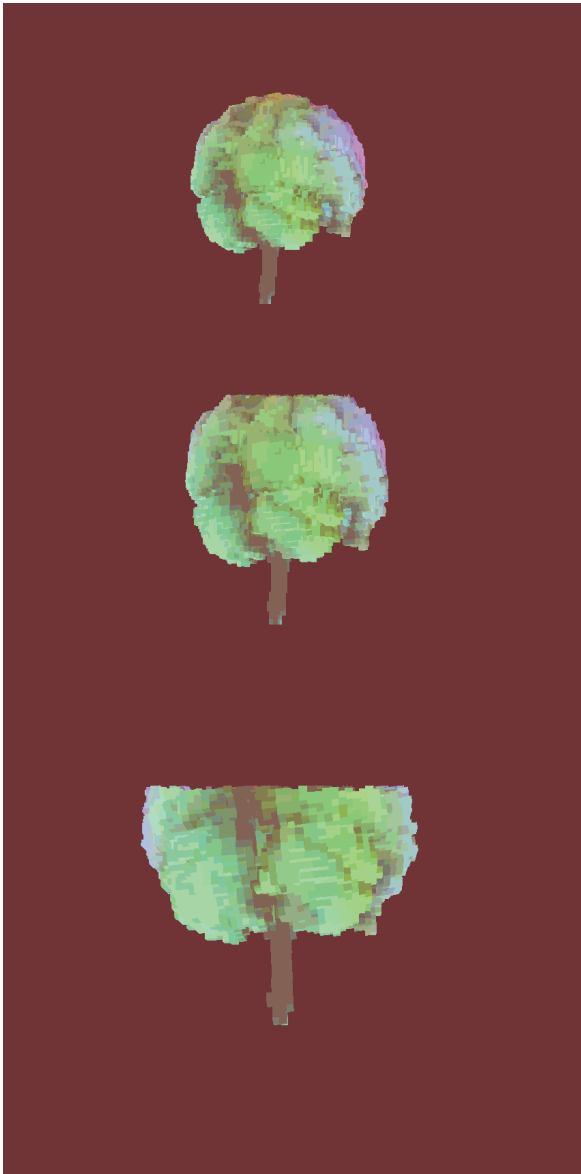


Figure 2 Ulysse's walking motion.



**Figure 3** An animation example.

#### 4. TACIT

The Tacit project [[16]] has been aimed at processing and understanding information from written reports of car accidents. Reports have been supplied by the MAIF insurance company and describe most often collisions between two vehicles.

Tacit combined two approaches. One was to build a deep model of semantics associated with the

reports. And the other approach was to use information extraction methods to skim some features of the accidents.

We used the deep semantic approach to generate and animate scenes in a virtual world from corresponding texts. In addition to language processing techniques, we also made an extensive use of knowledge on driving conditions in France.

## 4.1 MODELING THE SCENE

While the information extraction approach was applied to original running reports, the deep semantic modeling had to consider simplified texts. Car accident reports include many understatements, negations, or cosmetic descriptions that cannot be interpreted – at the present stage of our work – without guessing the state of mind of the driver. Consider for instance:

*Je roulais sur la partie droite de la chaussée quand un véhicule arrivant en face dans le virage a été complètement déporté. Serrant à droite au maximum, je n'ai pu éviter la voiture qui arrivait à grande vitesse. (Report A8)*

*I was driving on the right-hand side of the road when a vehicle coming in front of me in the bend skidded completely. Moving to the right of the lane as far as I could; I couldn't avoid the car that was coming very fast.*

where the collision is not even mentioned.

We needed a simplification to make this text explicit. It yields:

*Je roulais sur la chaussée, un véhicule arrivait en face dans le virage, je l'ai percuté dans le virage.*

*I was driving on the lane, a vehicle was coming in front of me in the bend, I bumped it in the bend.*

Even with this simplification, the text is merely a guide to re-create the scene and not a complete description: the author of the text gives the essential details only. The first clause of the text A8 for instance assumes that the driver follows the course of the road at a correct speed.

In order to reproduce this piece of reasoning, we used a processing architecture consisting of linguistic analyzer and a road domain model [[16], [17], [18], [19]]. Units identify entities and link them together using space and temporal relations. Linguistic processing splits sentences into a sequence of events. The result is a set of entities and the relations that link them. A major difference with Ulysse's was that Tacit involved an elaborate temporal modeling of events and actions [[20]].

Here are results we obtained with text A8 [[21]]:

*I was driving on the lane*

- Entities:  
v0: vehicle; s0: person; r0: road; e1: trajectory; im1, ip1, ie: intervals

- Time relations:  
contains(ip1, im1) &  
before(im1, ie)
- Relations:  
[ip1] driver(s0, v0)  
[ip1] within(v0, r0)

*a vehicle was coming in front of me in the bend,*

- Entities:  
v1: vehicle; s1: person; sr0: bend; e2: trajectory; im2, ip2, ie: intervals
- Time relations:  
contains(ip2, im2) &  
before(im2, ie) &  
simultaneous(im1, im2)
- Relations:  
[ip2] driver(s1, v1)  
[ip2] within(v1, rr0)  
[ip2] getting\_closer(v1, v0)  
[ip2] facing(v1, v0)

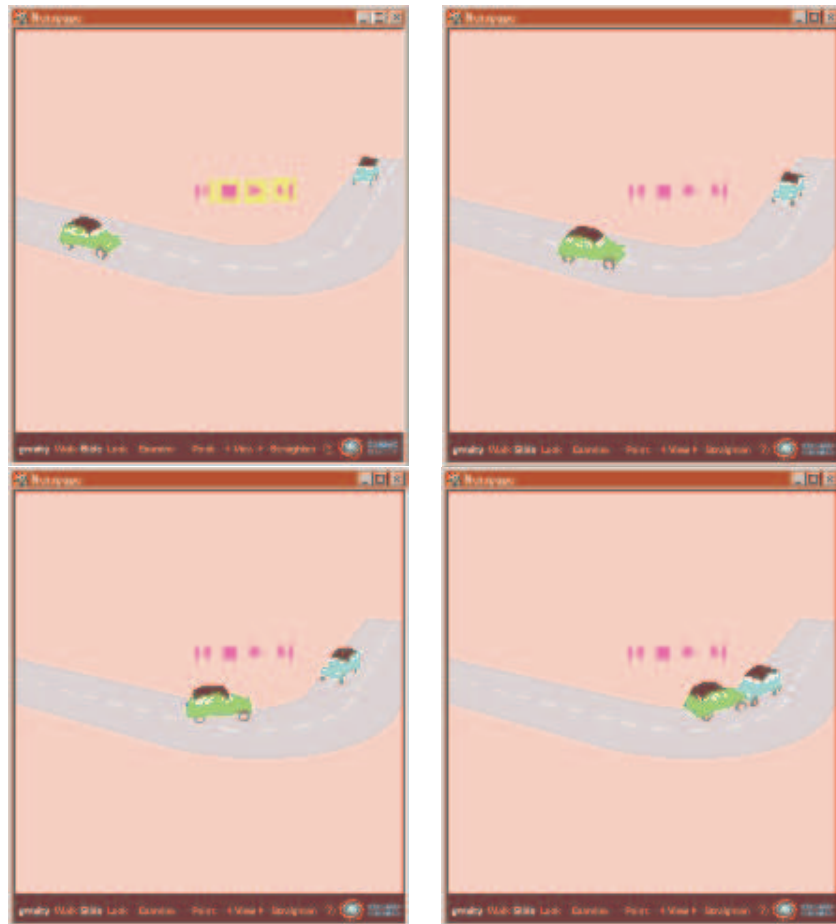
*I bumped it in the bend.*

- Entities:  
v1: vehicle; s0: person; e3: bump; im3, ip3, ie: intervals
- Time relations:  
contains(ip3, im3) &  
before(im3, ie) & before(im2, im3)

From the first sentence *I was driving on the lane*, we create a person s0 and his vehicle v0, a road r0, an event e1 corresponding to the movement of the driver. We also create time intervals (im, ie, ip) to locate the events in a chronological order and to determine whether they have finished or not.

We have relations specifying that the driver of the car v0 is s0, and that, during the process interval [ip1], the driver is on the road within(v0, r0).

These results are integrated in the scene construction unit. They are combined with knowledge of the domain in order to complete the description. The reasoning processes builds a chronology of events and the vehicles' trajectories. The description is translated into the VRML geometric format. It features a modeling of cars, roads, and road equipment. Animation considers the event list with their temporal length; the object list; their positions, speed and directions at the beginning and at the end of each process. It compiles them into VRML interpolators so that the scene can be animated. Figure 4 presents snapshots of the synthesized scene.



**Figure 4** Scene synthesis corresponding to text A8.

## 4.2 EMBEDDING A PLANNER

The temporal model we used in the previous example couldn't formalize well information relative to some driving events. For instance, it was impossible to encode the moving position of a vehicle relative to the road and at the same time, relatively to the position of another moving vehicle. We selected a couple of actions such as overtaking, or the behavior of vehicles at a junction and we modified the processing architecture. We implemented planners to reproduce these more complex events.

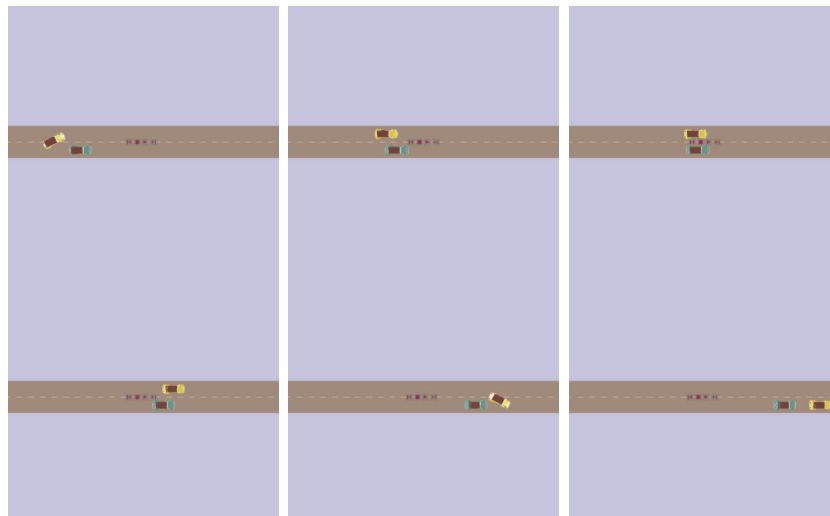
Representing an overtaking scene, for instance, requires the planner to split the overall action into a set of simpler ones. It also requires that the overtaking vehicle constantly monitors the position of the vehicle that is overtaken. We had to use a reactive planner because actions couldn't always be pre-computed and depended on the current state of the geometrical database.

From the Ulysse model where the planner is embedded in the user's avatar, we incorporated planners in vehicles. The planner is embedded in the tracking vehicle which is called the "actor". This vehicle updates its position relative to the vehicle in front: the "object". The planner applies rules to compute the successive positions and orientations of the vehicles within the time frame in which the action occurs.

The initial problem *overtake* is fractionated into sub-problems. It must satisfy an initial condition: “go faster than the other vehicle”. Then it must complete a sequence of actions: get closer to the leading vehicle; move to the left lane; go in front of the overtaken vehicle; finally return to right lane. Rules are triggered by comparing positions of the two vehicles. For instance, the actor changes lane when it reaches a given distance from the object. Knowing the positions and the entities of the scene, the system compiles the VRML interpolators that define the

## 5. VRML ULYSSE

Ulysse has recently undergone some changes. We rewrote it to adapt its architecture to Internet programming tools and languages, namely Java and the VRML programming interface. We also modified the parser that was quite slow eliminating the chart parser and replacing it by a more efficient algorithm. Although VRML Ulysse is not a complete port it



**Figure 5** Overtaking.

overall motion. Figure 5 shows snapshots of the synthesized overtaking.

Although the information extraction techniques can produce more substantial results, they are only able to extract templates from the reports. We believe that visualization and animation can make the text description easier to understand. They could help insurance analysts assess the likeliness of the report. We are presently considering them to check whether they could be useful to extract the sequence of events as the deep modeling did for simplified texts.

provides the programmer with a skeleton that is relatively easy to adapt to other tasks [[22]].

VRML Ulysse has three main components: a language engine in Prolog, the VRML world, and a Java applet to form the input interface and to ensure communication with the VRML world and the language engine. Both are linked through the External Authoring Interface. The Java applet is derived from the Script class that provides facilities to send and receive events from a VRML 2 world (Figure 6).



Master students from the university of Caen started using it to implement logical and cinematic definitions of French motion verbs. They worked on *sauter* (jump) and *courir dans* (run into) for which they implemented a model in Prolog that they could visualize with VRML Ulysse.

## 7. CONCLUSION

We have described three prototypes of conversational agents embedded within a virtual worlds. These prototypes accept verbal commands or descriptions from written texts. Ulysse conversational agent enables users to navigate into relatively complex virtual worlds. It also accepts orders to manipulate objects such as a virtual brain. We believe that a spoken or verbal interface can improve the interaction quality between a user and virtual worlds.

We have also described an agent to parse certain written descriptions of simultaneous actions of world entities. We have outlined techniques to synthesize and synchronize events from these texts. Although texts need to be simplified and the number of actions is still limited, we could animate the entities according to the description and simulate some of the accidents.

Finally, we have sketched the architecture of a new version of the Ulysse system in VRML, Prolog, and Java which allows the design of language applications in virtual worlds. It has been used to implement the cinematic definition of French motion

verbs. While there are theories in this domain, few are proven due to the lack of experimental devices. Experimental tools are central to the improvement or design of theories. We hope such a system enables the experimentation of theories on motion verbs making them implementable and provable. The prototype is available from the Internet.

In conclusion, we believe that the virtual reality and computational linguistics communities could have a fruitful cooperation. In spite of their different technical culture and history they could create paradigms to explore future interfaces and new ways of computing.

## REFERENCES

- [1] Benford, S. & Fahlen, L. "Viewpoints, Actionpoints and Spatial Frames for Collaborative User Interfaces". In *Proc. HCI'94, Glasgow*, August 1994.
- [2] Kosslyn S, *Ghosts in the Mind's Machine*, New York, Norton, 1983
- [3] Kosslyn S, *Image and Brain*, Cambridge, MA MIT Press, 1994.
- [4] Tufte E., *Visual Explanations, Images and Quantities, Evidence and Narrative*, Graphic Press, Cheshire Conn., 1997.
- [5] Godéreaux C., K. Diebel, P.O. El-Guedj, F. Revolta, & P. Nugues, An Interactive, Spoken Dialog Interface in Virtual Worlds, in:

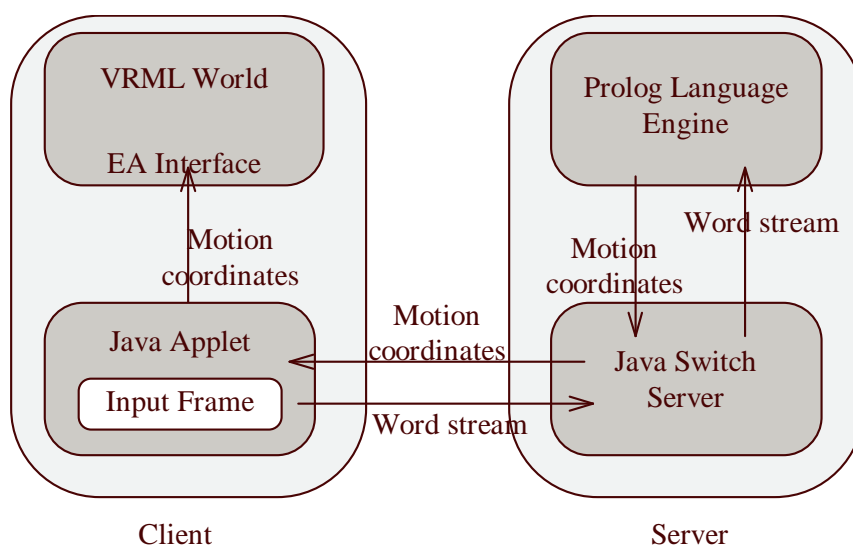


Figure 6 VRML Ulysse architecture.

- Linguistic Concepts and Methods in CSCW*, John H. Connolly and Lyn Pemberton (Eds), Chapter 13, Springer, London, 1996, pp. 177-200.
- [6] Gentner D. & Nielson J., The Anti-Mac Interface, *Communications of the ACM*, 39(8):70-82, 1996.
- [7] Godéreaux C., P.-O. El Guedj, F. Revolva, & P. Nugues, Un agent conversationnel pour naviguer dans les mondes virtuels, *Humankybernetik*, 37(1):39-51, 1996.
- [8] Nugues P., Godéreaux C., El Guedj PO, & Revolva F., A Conversational Agent to Navigate in Virtual Worlds, in: *Proceedings of the Eleventh Twente Workshop on Language Technology*, S. LuperFoy, A. Nijholt, and G. Veldhuijzen van Zanten eds., pp. 23-33, Universiteit Twente, Enschede, June 1996.
- [9] Bersot O., El Guedj P.-O., Godéreaux C., & Nugues P., A Conversational Agent to Help navigation and Collaboration in Virtual Worlds, *Virtual Reality*, 3(1):71-82, 1998.
- [10] Godéreaux C., PO. El Guedj, F. Revalta, & P. Nugues, Ulysse: An Interactive, Spoken Dialogue Interface to Navigate in Virtual Worlds, Lexical, syntactic, and semantic issues, in: *Virtual Worlds on the Internet*, J. Vince and R. Earnshaw (Eds), Chapter 4, IEEE Computer Society Press, Los Alamitos, 1999, pp. 53-70 & 308-312.
- [11] El Guedj PO & Nugues P, Analyse syntaxique combinant deux formalismes au sein d'un Chart hiérarchique, *Humankybernetik*, 38(3):135-144, 1997.
- [12] Fikes R. & Nilsson N. J. STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189-208, 1971.
- [13] Andersson, M., Carlsson, C., Hagsand, O., & Ståhl, O. DIVE, *The Distributed Interactive Virtual Environment*, Technical Reference, (Kista: Swedish Institute of Computer Science), 1994.
- [14] Ludwig M., A. Lenoir, & P. Nugues, A Conversational Agent to Navigate into MRI Brain Images, in: Proc. *Interfaces 97, 6th International Conference*, Montpellier, 28-30 mai 1997, pp. 207-210.
- [15] Damasio H., *Human Brain Anatomy in Computerized Images*, Oxford University Press, 1995.
- [16] Victorri B. at al. *Le projet TACIT, Traitement Automatique pour la Compréhension d'Informations Textuelles*, Rapport de fin de projet, Université de Caen, 1999, <http://elsap1.unicaen.fr/rapport.pdf>
- [17] Enjalbert P. & Victorri B., Du langage au modèle, *Traitement automatique des langues*, 35(1):37-64, 1994.
- [18] F. Pied, C. Poirier, P. Enjalbert & B. Victorri, From language to model, *Proc of ECAI-96*, 1996.
- [19] Poirier C., Mathet Y. Enjalbert P, La compositionnalité à l'épreuve des faits, à travers un projet de compréhension automatique de constats d'accident, *Traitement automatique des langues*, 39(1):99-129, 1998.
- [20] Gosselin L., *Sémantique de la temporalité en français*, Duculot, Louvain-la-Neuve, 1996.
- [21] Tabordet F., Pied F, & Nugues P, Scene Visualization and Animation from Texts in a Virtual Environment, *CC AI. The journal for the integrated study of artificial intelligence, cognitive science and applied epistemology. Special issue on visualization*, Guest editor: T.G. De Paepe, 15(4):339-349, 1999.
- [22] Baumard L., R. Bertholon, & M. Bittar, *Navigation dans un monde VRML*, Projet de fin d'études, ISMRA, Caen, 1998.

