Worlds Without Words

Ivan Bretan ivan@sics.se
Jussi Karlgren jussi@sics.se

Swedish Institute of Computer Science
Box 1263, S–164 28 Kista, Stockholm, Sweden.

Keywords: Natural Language Interaction, Virtual Reality.

November 23, 2005

Abstract

This paper points out a deficiency in virtual worlds as they are envisioned today. Several of the more important aspects of human communicative behavior are missing, since there is no way to accommodate natural linguistic interaction with and through the interface. We point out some explicit cases where natural language interaction would be useful in a VR framework, and identify some areas where natural language interpretation actually might benefit from being situated in a highly formalized environment.

Virtual Worlds As An Interactive Tool

Virtual worlds are often envisioned as front end tools for a user to interact with applications – a simulated world with an emphasis on the "physical" characteristics of the environment. In these cases, the focus in the research has been technically oriented towards how "physical" manipulation of objects can be done intuitively and robustly. The applications in mind have been, among others, simulations of actions and processes that would be difficult or expensive to organize in reality, and distributed applications, where agents – robots for instance – can be controlled from a distance.

Virtual worlds have also, lately, been proposed as a convenient tool to manage computer-mediated human communication and collaboration – as a medium for several human participants interacting with each other (see for instance publications by Fahlén *et al*, 1993). In these and similar cases, the human communication is assumed to be over shared multimedial channels, where the information between humans is not analyzed, but simply transmitted from one interlocutor to another.

Virtual reality research has concentrated on ways to make the environment manipulable for the user in the context of a room-space metaphor. As an example, a typical way to

"SELECT ALL THE GREY MARBLES."

Figure 1: Just point and click.

investigate the usefulness of the aural channel for the user is in terms of generating perceptual cues for user localization in the virtual space, and a typical way to use and share objects is to pick an object up, manipulate it, and hand it over or show it to someone also present in the same shared space.

In these settings, a user has the possibility of interacting with objects and manipulating them in various ways using gestures and other conventions of the physical world. These gestures are *intended* to be modeled on the gestures and other conventions of the physical world, to make the interaction in the virtual world intuitive. However, this world lacks something, and we will try to show, by some examples, what.

The Naming Of Things Is A Serious Matter

Virtual reality affords the user intuitively useful means of selecting and manipulating objects in the vicinity, much as gestures do in real life. Cognitive concepts like "this" and "that" are easily defined and formalized in virtual reality.

Human languages are by design a step beyond deixis or the simple acts of ostentation behind "this" and "that". They allow the user to refer to entities other than concrete objects,

"WHERE IS THAT NOTE ABOUT VIRTUAL REALITY I SENT CHRISTER YESTERDAY?"

Figure 2: Referring to things not quite here.

using arbitrary conventions: abstract concepts ("air", "knowledge", "algorithm"), actions ("running", "eating"), objects that are not present ("the stove I have at home"), objects that are no longer present ("my December salary"), objects that will be present ("Summer"), and objects that are impossible ("unicorn", "perpetuum mobile"), or objects with some specified property ("slow things").

None of this functionality is evident in what has been demonstrated as virtual reality tools today. The tools constrain their users to the *here* and *now*, even if "here" and "now" may be defined differently than in the physical reality.

In figure 1, the reference to "all the grey marbles" would be very difficult without the use of natural language.[1] The idea that someone might want to refer to grey marbles if they are represented as in the picture ought not to be surprising: the concept of the set of grey marbles, is not inherently complex.

In figure 2, the reference to an object which is not actually present will pose a difficulty, unless there is a way of referring to objects that are not visible by their temporal location or their content. Referring to "virtual reality", as in the example, without using human language will of course be a considerable challenge.

In general, rendering the domain of interaction in terms of physical objects is not always an appropriate representation for objects – many things are difficult to portray in a non-arbitrary way. Besides, in most domains we have to do with, there are simply too many referents we would like to use or potentially might like to use in interaction to be able to represent them in as any kind of accessible objects. Even with virtual space at our disposal, everything won't fit in!

Language and Action

The virtual world does not need to obey the laws of the physical. This is well established in virtual reality research and development. In the real world, language is a means to change the world, and in a virtual world the world will be easier to change. For instance, properties of the virtual world can be changed without physical manipulation, in contrast to our real world, using metaphorical action of some sort, or simple commands using some language. Of course, manipulation is always a conceivable interface to changing the world – but we may not always know what we want in terms of direct manipulation. Arbitrary manipulation of the physical reality is not always possible in the physical world. Language, however, knows no physical bounds. There is nothing stopping us from saying: "I will fly to the moon now".

[1] A similar point has been made by Dahlbäck and Höök, 1992.

"TAKE OFF ONE OF THE LEGS OF THE TABLE AND MAKE IT RED."

Figure 3: Changing the world by linguistic action.

"THAT TABLE IS AWFUL. MAKE IT NICE."

Figure 4: Changing the world by macros.

We can discuss something as simple as a virtual table. Unlike its physical relative, the virtual table can change its characteristics according to the preferences of the person interacting with it. However, there are not necessarily any natural metaphors that can be carried over to the VR scenario, to enable us to change the number of legs of the table, its color, shape etc... but see figure 3 for an example of how it easily can be specified in natural language. In fact, any means of specifying this sort of actions *other* than language will reduce the virtual reality to a 3-D direct manipulation interface, rather than a world.

As a further example of what language is good for, contrast the two examples in figures 3 and 4 where properties of natural language is made use of to compact a complex request. It is an illustration of the usefulness of natural language with its rapidly evolving conceptual structure – "macros" if you will.

The above leads us to the observation that language is not only about conveying information[2], it is a tool for acting in the world (a view emphasized by Winograd and Flores, 1986). Linguistic actions may not change the physical world as they could change the virtual ones, but they certainly change people. Entire communities are organized through linguistics actions, such as laws, decrees etc. but also through ordinary daily discourse.

In the virtual community, we can expect virtual members of this community to interact with human visitors. Whether we want to look upon these entities as being intelligent or not, being able to expose them to linguistic actions will further our goals of being in the virtual world in the first place. To discuss something less conventional than a table, real or virtual, we can imagine that we have a virtual assistant. We want to interact with it through orders, requests, acceptances, and rejections etc. The assistant can for instance be told to take us to a particular area in the virtual space, as in figure 5.

Die-hard VR believers seem to be striving for a state of interaction which in some sense

[2]In fact, as an experiment, the reader is invited to approximate how large a percentage of language use the reader personally uses for conveying *information*.

"TAKE ME TO BJÖRN'S OFFICE."

Figure 5: Navigation.

is more grounded and innate than all other means of using computers. Statements such as "[We are] on our own again, after the long mediation of top-down authored experience (...)"[3] bear witness to a conviction that we are returning to the *original* mode of being in the world. We must remeber that it is language that has made the world a meaningful place to exist and act in as humans, i.e., a world of communities (cf. Halliday, 1978). One fears that there is a somewhat naive belief among some VR advocates that the metaphors already invented for VR interaction, with some simple tuning and the addition of proper tactile feedback etc., are sufficient to create worlds in the true sense of the word.

Such a simplified standpoint can easily degenerate into the position that VR interaction does not require learning in the same sense as it is required for traditional interfaces, since the natural mechanisms of the user are taken advantage of. This is a familiar mistake. It has been made repeatedly in the natural language-processing community. Not until recent years has it been widely acknowledged that natural language interfaces as we know them today are hopelessly flawed, since they do not support "natural natural language", i.e. interaction where communicative intentions are not only conveyed through literal meaning of words, but through references to the surrounding context, relations to previous discourse, gestures, prosody, relying on mutual knowledge, and so on. In other words, language without the mechanisms that we make use of when interacting in the world is not natural language. Conversely, VR interaction without language does not take place in a natural world. Here we see an excellent opportunity to overcome some of the most fundamental weaknesses of these two paradigms of computing — through merging them.

References

Nils Dahlbäck and Kristina Höök "Can Cognitive Science Contribute to the Design of VR Applications?" 1992. *Proceedings of the 5th MultiG workshop.* Stockholm:SICS.

Lennart E. Fahlén, Charles Grant Brown, Olov Ståhl, Christer Carlsson. 1993. "A Space Based Model for User Interaction in Shared Synthetic Environments" *Proceedings of InterCHI*, Amsterdam:ACM.

M.A.K. Halliday. 1978. *Language as a social semiotic*, London: Edward Arnold.

Terry Winograd and Fernando Flores. 1986. *Understanding Computers and Cognition*, Addison-Wesley:New York.

[3]Brenda Laurel, WIRED 1:6