Natural Language Competent Robots

Stuart C. Shapiro
Department of Computer Science and Engineering
and Center for Cognitive Science
University at Buffalo, The State University of New York
Buffalo, NY 14260-2000
shapiro@cse.buffalo.edu

April 24, 2006

Why use natural language (NL) to communicate with computer systems? I have always thought the answer was obvious. Since NL is the most natural communication language for people to use [1], it would make computer systems easiest to use if they understood and spoke NL. I often get resistance to this opinion, but the resistance almost always takes the form of comparing current GUIs to current NL interfaces. Current NL interfaces do not work as well as current GUIs, because current computers do not fully understand NL. My feeling is that that just makes NL understanding (NLU) an important research topic. Indeed, not only is NLU in general an active research topic, so is the specific topic of NL interfaces to robots. For example, see the papers in the recent ACM Conference on Human-Robot Interaction [2]. Moreover, not only are people who describe themselves as NLU researchers working at making NL understandable to computers, the field of programming languages can be seen as a bottom-up attack on the same problem. Many advances in programming languages, as far back as the development of assembly languages, have been presented as allowing the programmer to express the problem to be solved in a high-level language that is natural for the human.

By “NL”, we could be referring to NL text or NL speech. Many NL researchers are working on text-based systems because they are interested in processing information from the world-wide-web, or other documents. Others are using text as an expedient until speech recognition becomes easier to use. (I’m in this group.) So, by “NL competent robot”, I will generally mean one that understands and generates speech.

What counts as a robot? By “robot”, I mean an embodied computer system that has sensor and effector organs. I am not, however, limiting the term to hardware-implemented robots that operate in the real world. I also include robots with simulated bodies, and sensor and effector organs that operate in simulated or virtual reality worlds. I also mean to include teleoperated robots as well as autonomous robots. A “teleoperated robot” is one that is controled by use of a console, joystick or other control device. I do not, however, want to argue that using prosthetic limbs or master-slave manipulators such as are used to manipulate objects in a toxic environment or to perform micro-surgery would necessarily benefit from NL communication.

There are at least two motivations for research into and development of NL competent robots. One is science. Since NL is the natural communication medium for people, researchers pursuing the AI goal of achieving a computational understanding of human intelligent behavior and creating devices that exhibit such behavior must include NL competent robots somewhere in their research agenda. (Though, maybe, not in their personal agenda.) The other motivation is pragmatic—the belief that robots will be more useful if they are competent in NL. Since I take the scientific motivation to be inarguable, I will concentrate on the pragmatics in the rest of this essay.

It might seem that controllers of teleoperated robots would not need to also use NL, given that they have more direct control devices, but it has been found experimentally that use of single word speech commands is a beneficial technique [3]. This is due both to the naturalness of speech, and to the ease of using speech when the operator's hands and eyes are occupied with the other controls. However, the use of a more complete speech-based sentence understanding system was not found to be beneficial, because it required the operator to pause between words, "and the loss in fluency was quite considerable" [3]. It was found that "the operator does not work comfortably if [the] recognition rate is under 90%" [3]. This is another example that when NL interfaces do not understand fluent speech, they do not compete well.

If speech is useful for robot controllers whose hands and eyes are occupied, it is even more useful for people who can't use their hands or eyes due to long-term disability or short-term injury. Speech interfaces for assistive devices is therefore an active area of research, and there were several papers on the topic at the aforementioned ACM Conference on Human-Robot Interaction [2].

While using speech as a command language is useful, even more useful is to use NL as a robot programming language. Any robot will have a repertoire of actions and recognizable domain features designed and/or programmed into it. A robust understanding system that allows a human controller to express these actions and features in NL would provide a useful speech command language. But now consider the situation if a human wants to ask the robot to perform a sequence of actions, or to do one of several actions depending on what it perceives in its world, or to repeat some action until it has accomplished some perceivable effect. As we know, the provision of NL constructs to express these control structures will make NL a full-fledged programming language. Some of these more complex procedures might be ad hoc procedures the controller wants done only once, but others might be more common. If the human can conceptualize these, possibly parameterized, procedures, express them in a way acceptable to the NLU system, and if the NLU system has a way for the human to say something like "The way to do X is to do Y," the NLU robot command language is equipped with all the constructs of a procedural programming language. This has been done to some extent by Lauria and colleagues in the context of a hardware robot that can be given directions for navigating in a small model city [4].

In order to use NL to instruct and command robots, the NLU system must translate NL inputs into some kind of robot acting program. That is, the meaning representation language used by the NLU system must be an autonomous-agent acting language. Unfortunately, researchers designing agent acting languages, and researchers designing representation languages for the (contextually-grounded) semantics of NL mostly operate independently of each other. A promising approach to building NL competent robots is to combine their efforts, making, for example, the representation used for the robot's understanding of a verb phrase be a construct in the robot's acting language that it can perform. In this way: the robot performs according to an imperative sentence by executing the structure representing its understanding of the command; by appropriately appending a representation of itself and of the current time to the representation of an action it has just performed, the robot has a belief that it has performed the action; giving that belief to its NL generation routine produces an NL report of what it has just done. This is the approach my colleagues, students, and I have been pursuing [5].

Given an NL programming language, it would also be important to have an NL development environment, and it would be most convenient if the robot, itself, participated in that. For example Lauria et al.'s navigation robot uses NL generation to ask for explanations of directions it doesn't understand [4].

It seems that people can't help but interpret speech in a social context, and attribute human characteristics to the speaker, even if they know that the speaker is a machine [1]. It is, therefore, important to take the social aspects of speech into account, so that "users will not simply talk *at* and listen *to* computers, nor will computers simply talk *at* and listen *to* users. Instead, people and computers will cooperateively *speak with* one another" [1, p. 184 italics in the original].

NL competent robots are not yet here, but progress is being made, and speech interfaces are becoming common at call centers and in automobiles [6]. Robotics researchers should be encouraged to include NL competence among the abilities of their robots, and NL researchers should be encouraged to consider robots as platforms for their work.

References

[1] Clifford Nass and Scott Brave. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. MIT Press, Cambridge, MA, 2005.

[2] ACM. 1st annual conference on human-robot interaction, 2006. `http://www.hri2006.org/`.

[3] M. Ferre, J. Macas-Guarasa, R. Aracil, and A. Barrientos. Voice command generation for teleoperated robot systems. In *7th IEEE International Workshop on Robot and Human Communication 1998 (RO-MAN'98)*, volume 2, pages 679–685, 1998.

[4] Stanislao Lauria, Guido Bugmann, Theocharis Kyriacou, Johan Bos, and Ewan Klein. Training personal robots using natural language instruction. *IEEE Intelligent Systems*, 16(5):38, 2001.

[5] Stuart C. Shapiro and Haythem O. Ismail. Anchoring in a grounded layered architecture with integrated reasoning. *Robotics and Autonomous Systems*, 43(2–3):97–108, May 2003.

[6] Elizabeth Millard. Talking computers just around the corner. *Sci-Tech Today*, March 14, 2006. `http://www.sci-tech-today.com/news/Talking-Computers-Around-the-Corner/%story.xhtml?story_id=0330030N4PO9`.