Natural Language Understanding and Action Control

Hozumi TANAKA
Tokyo Institute of Technology
April 18, 2002

Abstract

The natural language understanding (NLU) research environment has changed drastically in the past two decades. Better technologies in speech recognition, natural language processing and computer graphics are now available. We have obtained a huge amount of computing power under which research on computer graphics has made significant progresses in generating complex and realistic 3-D animated robots in cyber-space.

We are now in a good position to begin developing the next generation NLU system, combining three technologies: NLU, speech recognition and computer graphics. Such a prototype NLU system named Kairai was developed at our laboratory to carry out preliminary research on the next generation NLU. Kairai is certainly different from previous systems for its integration of the three technologies mentioned above. Although Kairai includes many inovative features, several problems hindering the building of a better next generation NLU system still remain. After giving a brief introduction of the Kairai system, we will conclude by outlining what problems ought to be solved in the future.

1 Introduction

We have been spending most of our time in physical space, but recently, due to the rapid progresses in computer science and network technologies, the time spent in cyber-space has been increasing. Thus, it is very important for us to make the cyber-space more comfortable and friendly. The interaction between human and information systems or software robots in cyber-space will be continually increasing.

However, the lack of the ability for a software robot to understand natural languages will make the cyber-space unattractive. As the natural language is the most natural and powerful communication means, we would like to use it as the communication medium in cyber-space as well as we do in the physical space. The interaction between human and software robot through natural language becomes natural even for persons inexperienced in cyber-space. For this reason natural language understanding (NLU) research remains very important.

In the past quarter of a century, a lot of NLU research has been carried out with many fruitful results. One of the most important NLU systems was SHRDLU developed by Winograd at MIT in the early 1970's [Winograd 1972]. This system was a kind of a software robot that worked in a toy block world simulated in cyber-space. However, rather than head, feet and hands, the robot was equipped with a small stick. It could understand English dialogue input from a user's terminal (no speech input) according to which it performed very simple tasks such as "Pick up a red block on the table" and "Put it in the green box". Its distinctive features were as follows.

SHRDLU combined NLU and task execution with results displayed on the terminal screen as an animation that imposes deeper analysis requirements on the part of NLU system. The system could also answer simple queries about the current state of the toy block world. It was enable to resolve anaphoric ambiguities, and build a plan to carry out a task specified by the input sentence. In the course of anaphora resolutions and a task planing, it carried out inferences by making reference to the current state of the cyber-space which it was contained in.

Although research on natural language processing and computer graphics was at a premature level at that time, SHRDLU demonstrated the promising future of NLU research. Many NLU projects followed, but they mostly ended without major progressive steps, since NLU remains to be one of the very difficult research areas. It is not exaggeration to say that an epoch-making NLU system that combines NLU and robot's actions has not been developed since SHRDLU.

The NLU research environment has changed drastically in the past two decades. Better technologies in speech recognition, natural language processing and computer graphics are now available. We have obtained a huge amount of computing power under which research on computer graphics has made significant progresses in generating complex and realistic 3-D animated robots in cyber-space [Badler 1993][Badler 1999].

Compared to hardware robots, which have severe mechanical limitations pertaining to movability, software robots have capabilities not only to carry out complex movements but also to change their facial expressions. It is possible for us to issue rather complex commands to a software robot. By combining at least three technologies: NLU, speech recognition and computer graphics, we are now in a good position to begin developing a next generation software robot that understands rather complex natural language expressions.

2 Next Generation NLU System

The question that needs to be answered is: What kind of research is necessary to build an intelligent software robot or the next generation NLU

system that understands natural language expressions. Consider the following dialogues:

1. Human: "Open the curtain covering the window a little."
2. Robot goes to the curtain, and grasps it by his/her hand and opens it.
3. Human: "A little bit more."
4. Robot opens the curtain a little bit more.
5. Human: "Too much."
6. Robot closes the curtain a little.
7. Human: "Air in the room is polluted."
8. Robot opens the window.

The first command issued by the Human makes the Robot create a plan to go to the curtain, grasp and open it. Such a plan is called a macro level plan. There are many ways to grasp and open the curtain. Robot has to select one of them to generate a micro level plan in order to carry out his/her actions. We can conclude that the Robot certainly understand natural language by watching the Robot's actions corresponding to what Human says. In other words, the Robot's actions, which are visualized in cyber-space as an animation, verify the NLU ability of the Robot. The visualized actions provide us with a NLU evaluation method more severe than that of Turing test, since the latter does not take account of visualized behavior of AI systems.

The second and third commands lack a verb in addition to a subject and an object. Robot has to augment these elliptical words by considering the context of a dialogue. With respect to the second command, Robot has to infer "open" as an appropriate elliptical verb, and then carries out the action "open." On the other hand, in the third command, "open" is also a correct elliptical verb, but Robot has to perform an opposite action "close" in this case. In other words, the Robot has to extract the intended actions from indirect speech act commands [Cohen 1990]. The second and third commands are related to the problem of vagueness, which were sometimes overlooked in the past NLU research.

The fourth command includes a typical indirect speech act that should be understood correctly for Robot to perform "open the door." Extracting true intentions in indirect speech act is one of the very difficult computational tasks.

In summary, the next generation NLU systems has to consider at least following items.

1. Resolution of anaphoric expressions.
2. Augmentation of ellipsis.
3. Extracting true intentions from a command.
   In additions to the above three items, it is necessary to include:
4. Combining technologies of speech recognition, NLU and computer graphics.
5. Handling ill-formed sentences that include fillers, additions, repairs and repetitions.

   The fourth item brings about the fifth item.

3 Kairai System

For the feasibility study on the next generation NLU system tightly combining three technologies: speech recognition, NLU and computer graphics, we developed a prototype NLU system called Kairai [Shinyama 2000] [Shinyama 2001] [Tanaka 2001]. Kairai system incorporates several 3-D software robots with which we can converse. It accepts voice inputs (spoken inputs), interprets them and performs the tasks specified in cyber-space.

The task executions are visible on a display screen as an animation. There are four software robots in Kairai system. In addition to three visible software robots: a horse, a chicken, and a snowman, a cameraman is also a software robot controlling his camera to give a different perspective of the cyber-space. The cameraman and his camera are invisible on the display screen. The camera handling is specified through commands such as "Go near the horse." In consequence, the figure of the horse is enlarged.

Kairai understand what we say in natural language, especially the words such as "left", "right", "in front of" and "behind" that indicate relative location in cyber-space. Typical actions performed by the (visible) software robots are "Push", "Go", and "Turn."

Figure 3 shows the outline of Kairai system, which is divided into three parts: speech recognition module, NLU module, and animation generation module. The speech recognition module transforms speech input into a sequence of words that become input to the NLU module that analyzes the input by using both a grammar and a dictionary and then extracts a meaning structure called a frame structure along with anaphora resolution and ellipsis handling. The latter two are together called a discourse process and refer to the context of past dialogues between human and Kairai. After a task plan is created by the NLU module, it is forwarded to the animation generation module to yield an animation on the display.

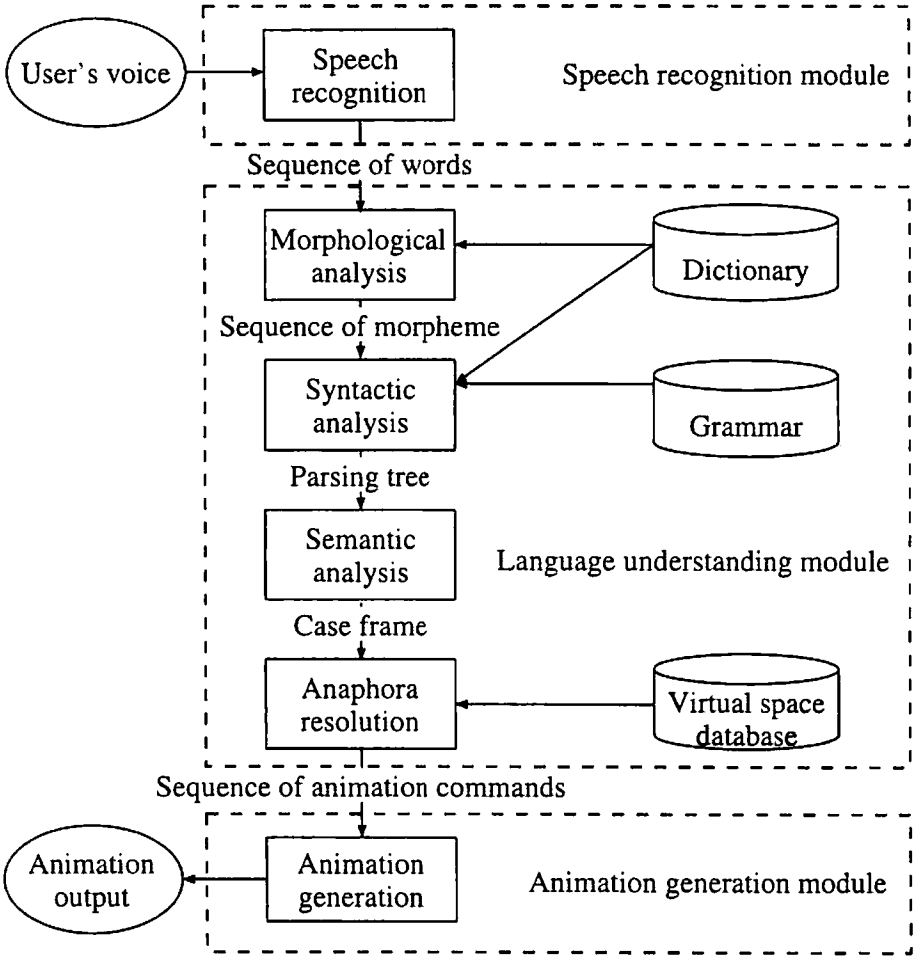

Figure 1: Outline of Kairai System

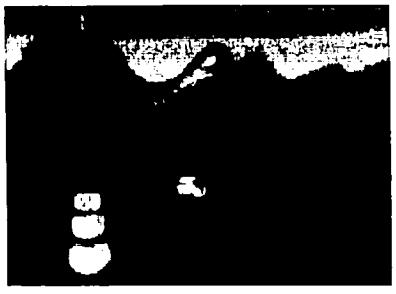

Figure 2: A Snapshot of Kairai

Figure 2 is a snapshot of animation generated by Kairai system. Readers can see three software robots in the cyber-space. According to the commands provided by the human, software robots including a cameraman move and perform appropriate actions in common space.

4 Problems in Kairai System

The experiences with respect to Kairai system which was developed as a prototype system to study the feasibility of the next generation NLU systems, brought realization of many remaining problems. The first and the most important problem is that Kairai was not really a multi-agent system composed of autonomous agents [Febler 1999] [Weiss 1999].

As each software robot (agent) in Kairai seems to carry out his action independently, Kairai system, at a glance, is a multi-agent system. However, Kairai system is not an actual multi-agent system. In addition to four software robots mentioned before, there is another special agent who knows everything in the cyber-space, receives and processes a sequence of words sent by the speech recognition module.

After accomplishing NLU tasks, the special agent decides which software robot should perform what kind of actions and then activates an appropriate software robot. As any software robot in Kairai system executes the task plan generated by the special agent, it is not really an autonomous agent in cyber-space. This is the reason why Kairai is not really a multi-agent system. The problem discussed here brings about another problem.

Since current software robots in Kairai are not autonomous, it is very difficult to conduct cooperative actions including several robots. Consider "gazing" [Rickel 2001], a simple cooperative action. In the current Kairai system, even though a software robot is conducting a task in the cyber-space, the other robots are not paying any attention to these actions. In human society, it is natural for a person to gaze at another one working near

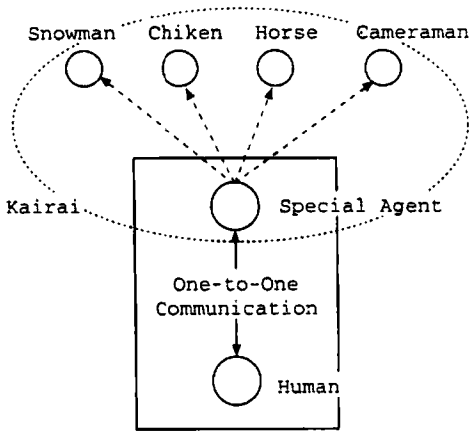

Figure 3: One-to-One Communication in Kairai

him/her. Due to the absence of autonomous robots in current version of Kairai, it is impossible for any robot to directly communicate with another. Problems of gazing as well as the other cooperative actions can be solved by introducing autonomous robots and one-to-many conversation mode in cyber space. The latter will be discussed more in the next section.

Currently, Kairai does not deal with paralinguistic phenomena including intonation in speech, gazing, facial expressions, body actions including hand gestures. Facial expressions are related to emotional actions. As paralinguistic phenomena play important role in communication, we would like to account for paralinguistic phenomena as one of challenging research topics. Fortunately, compared to hardware robots, software robots emulate paralinguistic behavior much easier since they do not have any mechanical limitations.

5 One-to-Many Conversation

In addition to the items mentioned in the section 2, NLU systems should deal with a one-to-many conversation in addition to one-to-one conversation. One-to-many conversation makes sense in the multi-agent environment, since in the one-to-one conversation it is easy to decide who the intended listener is. On the contrary, in the one-to-many conversation, as there are many potential listeners, it is difficult to decide to whom a command issued by a speaker is intended. Usually, the listener is mentioned explicitly in the first dialogue, but that he/she is not mentioned in the rest of the dialogue. Confusion can happen among agents if each agent is unable

Figure 4: One-to-One Conversation

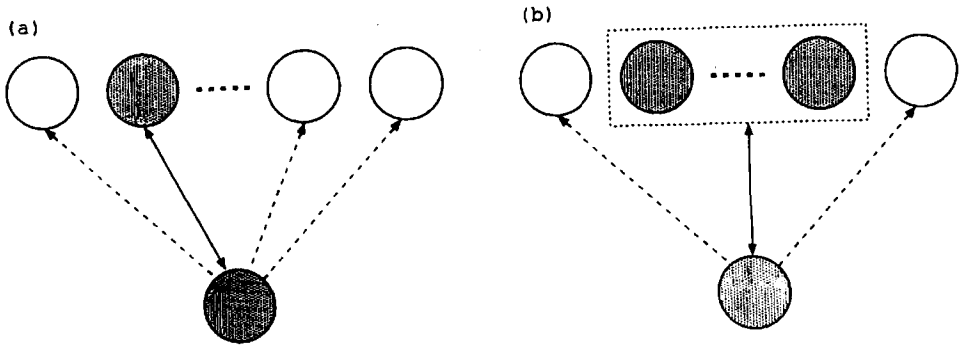

Figure 5: One-to-Many Conversation

to recognize who is the actual intended agent who needs to perform some tasks according to the command given by a speaker. The problem takes place when a subject or an object does not appear in a command due to ellipses.

To understand the above problems clearly, consider the following conversation in a multi-agent environment.

1. Human: "Hey, Robot A, I am throwing a red ball."
2. Robot A looks at Human.
3. Human: "Catch it."
4. Robot A begins the action to catch the red ball.

Note that the above Human command lacks a subject, but Robot A has to perform the action "catch" along with resolving the anaphora expression "it" that should correctly be identified as the red ball. The other robots should not perform the action "catch" even though they can hear "Catch it" command.

It seems obvious that each software robot should be autonomous in the multi-agent environment and have the ability to control his behaviour in his own right.

6 Conclusions

After reviewing an NLU system in the past, we pointed out several important issues concerning the building of the next generation NLU system. Kairai, a system which was developed at Tokyo Institute of Technology, played a key role in exemplifying the issues mentioned in the preceding sections.

Even though the system under consideration is composed of a set of software robots, the research results are applicable to any future multi-agent system consisting of hardware robots. We have also discussed the importance of paralinguistic phenomena in addition to emphasizing the need for better algorithms for anaphora resolution and ellipsis handling. Additionaly, dealing with ill-formed sentences which frequently occur in spoken language is also an important issue requiring attention.

The next generation NLU system ought to be a multi-agent system, with a wide array of application areas such as:

1. Entertainment
2. Hepler Robots (medical and in-home use)
3. Vitual Space

4. Electrical Appliances.

Let us expand on the final item. The future electrical appliances will be equipped with "ears" for listening to user's commands and will need to process these commands to execute them similar to current software robots. In such circumstances, the research on both multi-agent system and one-to-many conversation system will become increasingly important.

7 References

[Badler 1993] Badler, N. I., Phillips, C. B. and Webber, B. L.: Simulating Humans--Computer Graphics Animation and Control, Oxford University Press, 1993.

[Badler 1999] Badler, N. I., Palmer, M. S. and Bindiganavale, R.: Animation Control for Real-Time Visual Humans, Comm. of the ACM, pp. 65-73, 1999.

[Cohen 1990] Cohen, P. R., Morgan, J. and Pollack M. E. eds.: Intentions in Communication, The MIT Press, 1990.

[Febler 1999] Febler, J.: Multi-Agent Systems--An Introduction to Distributed Artificial Intelligence, Addison-Wesley Longman, 1999.

[Rickel 2001] Rickel, J.: Intelligent Virtual Agents for Education and Training: Opportunities and Challenges, in Angelica de Antonio, Ruth Aylett and Daniel Ballin (Eds.), Intelligent Virtual Agents--Third International Workshop, IVA 2001 Madrid, Spain, September 2001 Proceedings, Springer, pp. 15-22, 2001.

[Shinyama 2000] Shinyama, Y., Tokunaga, T. and Tanaka, H.: ''Kairai''--Software Robots Understanding Natural Language, 3rd Workshop on Human Computer Conversation, pp.158-163, 2000.

[Shinyama 2001] Shinyama, Y., Tokunaga, K. and Tanaka, H.: A Software Robot Kairai that Understands Natural Language (in Japanese), Journal of JIPS, vol.42, no.6, pp.1359-1367, 2001.

[Tanaka 2001] Tanaka, H.: Language Understanding and Action Control, Kaken Report (in Japanese), Ministry of Education and Science, Mar. 2001.

[Tanaka 2002] Tanaka, H.: Language Understanding and Action Control,

Kaken Report (in Japanese), Ministry of Education and Science, Mar. 2000.

[Weiss 1999] Weiss, G. ed.: Multiagent Systems, The MIT Press, 1999.

[Winograd 1972] Winograd, T.: Understanding Natural Language, Academic Press, 1972.