Exploring Interaction between Humans and Robots with Conversation Analysis

The Ido-Robo Project seeks to place the understanding of human beings' daily interactions, such as conversations and gestures, in informatics. It applies interdisciplinary efforts to discover knowledge that can be exploited in future robot design. In this respect it has deep connections with the field of conversational analysis (CA), the specialty of Professor Aug Nishizaka, a sociologist. Assistant Professor Mayumi Bono, who has a background in the humanities and is also engaged in CA, discusses the role played by sociology in human-robot interaction with Professor Nishizaka.

Investigating the Structure of Utterances

Bono Traditionally, research of conversational robots has proceeded within a framework that considers a robot as being able to talk if a speech-recognition system is created and incorporated into the robot. In contrast, the Ido-Robo Project is pursuing knowledge such as the "structure" of interaction with two or more people. In other words, I hope to offer another perspective for social robotics research by exploring the structure of human social exchange.

One pillar of this research is the study of "robot-human theater." I'm exploring the relationships between fine body movements and gaze movements and conversation as I observe director Oriza Hirata's production and collect a variety of data. As part of this study, I'm focusing on the start of conversations and communication. I think the methodology of conversational analysis (CA) can be used to analyze the flow of lines of script as they are repeated under Oriza's direction during rehearsals.

Nishizaka How fascinating. In CA, turn-taking rules on who speaks next in a conversation are formulated in such a way that at first glance, it looks as if they can be converted directly into a computer program. Because of this, at one time AI researchers thought they could use the rules to create conversational robots. But the results didn't turn out so well. This is because even if the rule can be formalized, actual conversations depend all too much on context.

Turn-taking rules mean prioritizing between two techniques that select the next speaker. Either the current speaker chooses the next speaker, or the next speaker begins to speak on his or her own. The first technique is composed of two elements. One, the current speaker addresses utterances to a particular person. Second, the speaker makes utterances that strongly elicit a particular behavior. For example, if the speaker addresses a question to a specific person, that person becomes the next speaker and is given a strong prompt to reply.

Bono But, in actual cases, what happens doesn't always work according to theory, right?

Nishizaka Right. Suppose the sentence that the speaker says is a question in Japanese. Even so, such a sentence doesn't always end explicitly with the sentence-ending particle "-ka" to indicate that it is a question. What's more, it is not always clear to whom a question is addressed. If I ask in Japanese, "Bono-san no shusshin wa doko?" ("Where's Bono-san from?"), if you are in front of me, then the question is addressed to you. But if you aren't there, then the same question can be directed to someone who knows you. In other words, it all depends on context. Formalizing this structure of "utterance design" to allow robots to have conversations is extremely difficult. To put it another way, this shows how flexible humans are in their adaptability to context. I think what is deeply intriguing about the Ido-Robo project is that it is not an attempt to create a robot that can converse like a human being, but rather it seeks to bring about perceptional changes by placing something in the likeness of humans in our midst to study how we converse.

Mayumi Bono
Assistant Professor, Digital Content and Media Sciences Research Division, NII
Assistant Professor, Department of Informatics, School of Multidisciplinary Sciences
The Graduate University for Advanced Studies
Principal Investigator, Ido-Robo Project

Robot-ness, Human-ness

Bono In the first place, engineers segregate conversations from context all too easily when they design robots. That's why I started the Ido-Robo project. I want to incorporate the complexity of conversation into the design of social robots a little more. Even at a performance of the robot-human theater, there are actually a variety of things that can start a conversation, such as the actors or robots' casual nodding to each other, making eye contact, and moving closer together. For example, imagine a scene where a robot that gathers data in a lab is giving a tour of a facility. As it is having a conversation with a character, another character walks by. The passerby's gaze is received, and a complex interchange of gazes between the three characters takes place. What's more, added into all this are Oriza's detailed directions, such as when to look at a character, how to return a gaze, and the robot's nod to the passerby. So, even when the passerby is a silent third party, his presence increases the interactions. The relationships clearly change when a three-person conversation takes place.

Nishizaka I'm excited to see how robot-ness is staged in the plays. After all, robots and humans are different in their

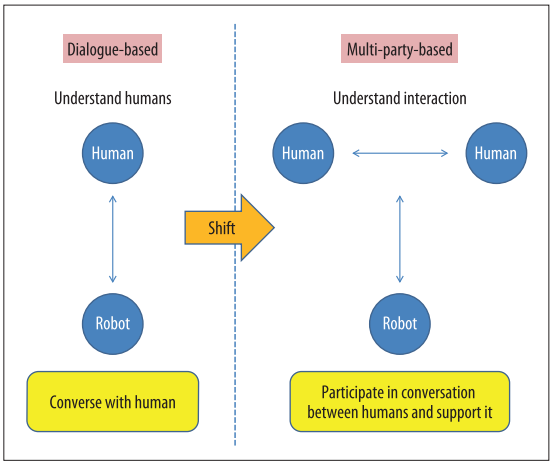

Figure 1 Core of Ido-Robo Project

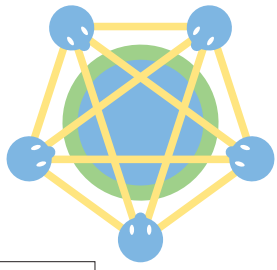

Figure 2 Director Oriza Hirata stages an exchange between robots and customers at a shopping mall to examine the interaction between humans and robots in the future. There are nearly 50 patterns of interaction. Here, human actors play both human and robot roles. (Photos courtesy of the Advanced Telecommunications Research Institute International [ATR], taken on May 14-15, 2013.)

essence. Oriza's project can be seen as an attempt to reveal the difference between humans and robots by showing just how humans and robots that behave like humans differ.

Giving the Insights of Conversation Analysis **Back to Society**

Nishizaka CA is originally a research area under sociology. It is a methodology to observe and analyze the interaction between human beings called "conversation." The concept of analyzing conversations arose as personal audio recorders became available in the early 60s. This made it easy to record and play back voice recordings. Next, the invention of compact video cameras made it convenient to record real-life scenes. So sociologists started paying attention to how visual information such as gazes and gestures function as resources to realize conversations. They found that while they could systematically extract elements that are in play in the exchange of words, formalizing conversation resources besides talk was difficult, because those other resources were too analog.

Bono Non-talk elements are still difficult to formalize, aren't they? But when it comes to talk, long years of CA research in that area have brought beneficial results to society. For example, I've heard that analyses of calls to the police and fire departments have proved useful as materials in court. Professor Nishizaka, you recently analyzed conversations by survivors of the Great East Japan Earthquake at footbaths in shelters in Fukushima, and wrote about your findings in a book, didn't you?

Nishizaka In the aftermath of the disaster, the shelters were still in chaos. The evacuees frequently didn't know what they themselves needed. One helpful piece of knowledge gained from the experiences of the Great Hanshin Earthquake and the Chūetsu Earthquakes was to set up footbaths. When the evacuees used them, they naturally talked about the things that were troubling them. So, I thought that by applying CA, we could explicate the structure of communication at the footbaths. I thought we could contribute to volunteering at the shelters this way.

Bono What did you find?

Nishizaka The evacuees may find it easier to tell their experiences thanks to hand massages and footbaths. But more than that, we understood how important the structure of the interaction was. Even if the conversation itself stops, the interaction continues. In other words, massaging is the base that supports interaction. It actually serves as the foundation to the very end of the interaction, but doesn't hinder conversation that is built on top of it. So, as the evacuees receive hand massages at the footbaths from volunteers, it's not unusual that they naturally tell about their experiences, even if the volunteers don't ask. This finding can show the way communication takes place in places like shelters and temporary housing if similar structures are present, even if the footbaths themselves are not there.

Bono New interactions arise as people do something while conversing—this story relates to the Ido-Robo Project, doesn't it?

Nishizaka The structures of human interaction are revealed by placing robots in the middle of interactions. That is precisely the appeal of the Ido-Robo Project. I look forward to this research's progress.

(Written by Shigeyuki Ohara)

Aug Nishizaka
Professor, Department of Sociology
Meiji Gakuin University