Machine Translation in the Mobile and Wearable Age

Nigel Ward
University of Tokyo
Mechano-Informatics, School of Information Science and Technology
7-3-1 Hongo, Bunkyo-ku, 113-8656 Japan
nigel@sanpo.t.u-tokyo.ac.jp

Abstract

The growth of mobile and wearable devices brings new challenges and opportunities for computational support for cross-language communication. Current research efforts, however, are addressing only a small subset of the potential uses of machine translation and related technologies. This paper discusses some issues in the choice of appropriate research aims and usage scenarios for mobile and wearable communication aids, and suggests that more effort be given to modeling the interpersonal, pragmatic, and situation-dependent aspects of language and communication.

1 Idealistic and Incremental Research Aims

Some of the most exciting prospects for machine translation involve mobile and wearable applications. Work in this area is largely driven by usage scenarios: that is, visions of the sort of system that should be built. Today's common usage scenarios arise, it seems, from two sorts of consideration.

The first source of scenarios is science fiction and fairytales. A common feature in many novels and movies is some sort of magical translator, such as Douglas Adam's original Babblefish. This seems to be the inspiration for much of the interpreting telephony work in the CSTAR consortia. The nice thing about this as a research goal is that the aim is perfect, translation, transparent to users. This is clearly attractive, and understandable to the man on the street, and to taxpayers and funders. Moreover, since the goal is nothing short of perfection, work on any and all sub-problems in machine translation is justified under this vision.

However, this research strategy allows no place to think about how people might actually use systems. Since the assumption is that the (eventual) product will be a flawless, transparent translation device, there is no reason to think about the user interface issues at all. It is as if a prosthetic module for a new language were to be magically attached to the brain, without any need for the user to be aware of or to pay conscious attention to the task of using the device.

A second source of usage scenarios is today's technology. The business meeting and meeting scheduling scenarios of Verbmobil project, for example, were chosen explicitly since they were likely to be achievable with (some aggressive improvements to) existing technology. "Achievable" here does not mean the ability to create a useful or even

[0]With thanks to Jani Patokallio and the Casio Foundation

usable system, but the possibility of reporting fairly high accuracy figures, and of producing demonstrations that seem plausible to cooperative users and to third-party observers. In other words, these scenarios were chosen not to provide a challenge leading to new ideas, new priorities, or new approaches; but in order to showcase existing technologies, and provide a safe arena in which incremental improvements would seem like real progress.

This research strategy is advantageous in that it provides clear goals that advance the state of the art, but it is not always clear that the advances are in the direction where they are most needed. Here too, there is an unexamined hope that the research results will eventually turn out to be useful, with no specific justification for that faith. (Of course, some researchers in these projects are doing basic research, of value for its own sake, but most of the effort is devoted to technology development and application.)

These two inspirations have given birth to the research field called "speech-to-speech translation". As occasionally noted in the research literature, this is very different from "interpretation", but the significance of the gap often is forgotten. Certainly there is a generation of students who now believe that the interpretation problem has been solved, because of system demonstrations they have seen on television.

The differences between interpretation and translation are too numerous to discuss in depth, but a consideration of the real-time nature of interpretation suggests the magnitude of the problems not yet addressed. Issues related to real-time communication include turn-taking, differences in word order, and the synchronization of non-verbal behaviors with language. While these issues have recently received some research attention, it is fair to say that that no-one has any conception of how to deal with them in the context of the common usage scenarios mentioned above.

2 A Focus on Usability

There is, however, another way to propose a usage scenario, to focus on user needs. Most generally, this means looking for ways in which technology can and should be used to help alleviate the problems which arise from the existence of different languages in the world. That is, looking at the ways which people achieve cross-language communication today, we can look for opportunities to use machine translation algorithms and devices to assist them.

There are two ways to do this. The first is to examine what human interpreters do. Most research on interpretation, it seems, focuses on simultaneous conference interpretation, however this is largely a uni-directional, non-interactive activity, and is such atypical. What most interpreters do is facilitate communication between two people in real time, an altogether much more involved task (Roy 2000). In such contexts, the interpreter is not invisible, but often takes an active part as a participant in the the conversation.

The second way is to look at what people *without* interpreters do, as done in the following two sections.

3 A Different Usage Scenario

Imagine that Beth has just arrived at the Mahale airport. Leaving the airport, she loads a Rutungu interaction module into her wearable computer, and goes over to the train ticket counter.

While waiting in line she clicks thru to reach the train-station phrase menu, then reviews the options available.

When she reaches the agent, she smiles and clicks to launch the standard introduction, the Rutungu equivalent of

> "Hello. I do not speak Rutungu, so I will talk through this interpreting device. Is this OK?"

The ticket agent looks surprised for a second, then looks troubled. As he says something curt with an abrupt hand gesture, Beth pushes forward her map and launches the next phrase.

> "Thank you. I would like a round trip ticket to here"

says the machine, as Beth points to the city she has circled. The agent again says something curt and turns to his listing. Beth waits. In a moment he comes back and says something she which interprets as a question. Looking confused and apologetic, she launches:

> "I'm sorry, my machine can only translate one way. Is a ticket available?"

The agent looks annoyed, then has an idea. He holds his fingers to his lips and makes a puffing motion, then raises his eyebrows and points at Beth. Beth clicks through two menus and launches:

> "No-smoking, if available, please"

The agent says something as he turns away, but Beth decides not to pursue it; and in a moment he comes back with a ticket. He points to the price, and as Beth finishes paying, she launches:

> "Which platform does the train leave from?"

The agent answers while gesturing a path with his finger.

> "Could you please write that down for me?"

The agent scribbles the number 6 on the back of the ticket. Beth takes it, smiling her thanks as her machine says.

> "Thank you very much."

Studying the ticket, Beth notices she has 10 minutes before the train leaves. She finds what looks like a noodle stand in the corner, calls up the restaurant script and scans the menus to pre-load the phrases she will need to order noodles in a meat-less broth, and then walks over to have lunch.

4 Issues in Wearable Translation

In normal human-to-human conversation, non-verbal cues are important. Gestures indicate where things are, facial expressions indicate degree of understanding, posture indicates attitude, tone of voice indicates invitation or query or request, and so on. In cross-language communication also, it is a common experience that gestures can get you a long way, even if no words are understood. Thus we believe the use of a wearable device with a heads-up display may allow a person to communicate with someone in another language better than they could by using a translation device which they have to look down to use. Moreover, given the limitations of speech recognition technology, there is no possibility of building a translation device capable of general bi-directional translation for at least a decade or two. Given this, it is essential to engage the non-verbal communication skills of the participants.

Thus the proposal is to only partially automate the communication process: to exploit the strengths of both man and machine to produce a hybrid solution to the problem of communicating across a language barrier. This was seen in the above scenario, where there was a division of labor between Beth and her translation device. The device actually output the sentences, but Beth was responsible for everything else: deciding when to launch each utterance, using smiles and gestures to elaborate on the utterances, and interpreting the agent's utterances, gestures and actions.

A device as described above would, of course, not be fully general. In particular, it would only work for dialogs in airports, train stations, restaurants, shops, hotels, etc., where the simplicity of these dialogs means that the translation does not have to be bi-directional, since the probable responses of the native are few enough in number that the user can generally classify a response based on non-verbal information alone, and where the number of things the traveler will need to say are finite.

This proposal differs interestingly from the other approaches to wearable translators.

The Diplomat/Tongues project at CMU (Frederking et al. to appear; Frederking et al. 1997) was designed to "explore the feasibility of creating rapid-deployment, wearable bi-directional speech systems". The focus of the effort was on the technical feasibility of speech-to-speech translation, specifically the three component technologies — machine translation, speech recognition, and speech synthesis. Questions of usability, have not yet, it seems been seriously considered. Rather, the inadequacy of current speech and language technologies led the system designers to cast the users in a supporting role: the users are enlisted in the task of preventing errorful translations (although this sort of "interactive editing" has not been well accepted in other machine translation applications (Ward & Jurafsky 2000)). Performing this role requires both users to have access to a GUI running on a laptop, requiring a style of interaction which diverges from the more common visions of wearable computer use.

Recently the LingWear "mobile tourist information system" at Karlsruhe has been reported as including a "translation module" with spoken output and perhaps input (Fuegen et al. 2001), but no details on the design have yet been published.

Hand-held translation aids are another area of research activity. Descended from the venerable phrase book (Berlitz, etc), and electronic equivalents such as the Canon

WordTank, systems have recently appeared with speech output. More recently, a hand-held portable electronic dictionary with speech input was proposed by (Obuchi et al. 1999).

Elsewhere we describe a system implemented based on the above considerations (Patokallio & Ward 2001a; Patokallio & Ward 2001b). In a study with 11 non-Japanese-speaking tourists, 10 found the system usable and were able to obtain directions from passers-by with it, although some problems remain.

5 Implications

The system described above is, in a sense, a perfect baseline for studies in wearable translation, because currently it does no translation whatsoever.

The sort of translation functionality that is needed here is clearly different from the sort of translation which most research to date has focused on. The need here is not for compositional, general-purpose translation, but rather on translation that involves the interpersonal, pragmatic, and situation-dependent aspects of language and communication, and on translation that is done in the service of a hybrid solution to the problem of communicating across a language barrier.

In general, recent years have seen more diversity in the usage scenarios proposed for machine translation (Lazzari 2000), but there is still a need for further explorations, especially in mobile and wearable applications. These should be developed, not purely idealistically nor based entirely on incremental technology-driven considerations, but with attention up-front to issues of usability.

References

Frederking, Robert, Alexander Rudnicky & Christopher Hogan: 1997, 'Interactive speech translation in the Diplomat project', in *Spoken Language Translation Workshop*, ACL.

Frederking, Robert, Alexander Rudnicky & Christopher Hogan: to appear, 'Interactive speech translation in the Diplomat project', *Machine Translation*.

Fuegen, Christian, Martin Westphal, Mike Schneider, Tanja Schultz & Alex Waibel: 2001, 'LingWear: A Mobile Tourist Information System', in *HLT 2001 preliminary proceedings*, pp. 373–377.

Lazzari, Gianni: 2000, 'Spoken translation: Challenges and opportunities', in *International Conference on Spoken Language Processing*, pp. 430–433.

Obuchi, Yasunari, Atsuko Koizumi, Yoshinori Kitahara, Jun'ichi Matsuda & Toshihisa Tsukada: 1999, 'Portable speech interpreter which has voice input and sophisticated correction functions', in *Eurospeech99*, ESCA, pp. 2023–2026.

Patokallio, Jani & Nigel Ward: 2001a, 'A design for a wearable translation device', *Human Interface*, **3**(3): 5–10.

Patokallio, Jani & Nigel Ward: 2001b, 'A wearable cross-language communication aid', in *International Symposium on Wearable Computing*, pp. 176–177.

Roy, Cynthia B.: 2000, *Interpreting as a Discourse Process*, Oxford University Press.

Ward, Nigel & Dan Jurafsky: 2000, 'Machine translation', in Daniel Jurafsky & James H. Martin, eds., *Speech and Language Processing*, Prentice-Hall, pp. 720–751.