Real-time Multilingual Communication by means of Prestored Conversational Units

Norman Alm¹, Mamoru Iwabuchi², Peter N. Andreasen¹, Kenryu Nakamura³, and Iain R. Murray¹

¹Department of Applied Computing, University of Dundee, Dundee, Scotland
²Faculty of Education, Hiroshima University, Hiroshima, Japan
³Department of Special Education, Kagawa University, Takamatsu, Japan

nalm@computing.dundee.ac.uk

Abstract

A computer mediated communication system has been developed which can offer real time multilingual communication, as long as users stay within the boundaries of prestored conversational units. The system was designed originally to give non-speaking people a multi-lingual capability. It is based on research into conversational modelling and utterance prediction, making use of prestored material. However, the system could also be used by people whose only communication disadvantage is not being able to speak a foreign language. The prototype consists of a large store of conversational texts, each one produced in several languages, along with a model of conversation which allows the system to link the items together into appropriate sequences. Both the non-speaking person and the communication partner use the communicator in their dialogue. In comparison with a multi-lingual phrase book, the system helped users to have more natural conversation, and to take more control of the interaction. This project is an interesting example of the way in which systems developed for people with severe disabilities can often have useful general applications.

1. Automatic translation technology

Automatic translation and speech recognition systems have developed considerably in the last decade in particular. However both of these technologies fall short of being able to provide real-time, completely accurate interlanguage translation. The current situation is that until unrestricted speech recognition in noisy environments is a reality, automatic translation systems will need to rely on text which the user must somehow produce. Even when working with text, however, available translation technology is still not accurate enough to be effective in a fast-paced interactive situation.

Various software packages for machine translation are available and are now widely used for the rough translation of text, with their effectiveness dependent on the subject matter [1]. The following examples of translation generated by typical systems illustrate the problem of translation inaccuracy. The first example shows some English text automatically translated into Spanish. The second example shows the same text automatically translated into Japanese. (SYSTRAN PROFessional and LogoVista E to J were used for the English-Spanish and English-Japanese translation, respectively. SYSTRAN PROFessional is a trademark of SYSTRAN Software, Inc. LogoVista E to J is a trademark of LogoVista Co. Ltd.)

Original text:
Sorry, your wheelchair must be checked in. We will lend you one of ours.

Example 1
Spanish translation:
Apesadumbrado, su sillon de ruedas se debe llegar. Le prestaremos uno el nuestros.

Which literally means:
Grieved, your armchair of wheels must arrive itself. We will lend one the ours.

Example 2
Japanese translation:
Sumimasen, anata no kuruma-isu wa touchaku o kiroku sare nakute wa naranai.
Wareware wa anata ni wareware no no 1-tsu o kasudearou.

Which literally means:
Sorry, the arrival of your wheelchair must be recorded. We will lend one of ours to you.

The degree of error here is not unusual. Where the system is being used to translate a webpage, or a document, the user can either settle for a roughly accurate translation, or use what is produced as a first pass at the material, editing it into accurate form. In a
spoken dialogue, however, the unedited version could cause confusion.

An additional problem for automatic translation of spoken dialogues as has been said, is transforming the spoken language accurately into text. Speech recognition technology has made considerable advances, but producing perfect text from the free-flowing speech of any speaker in a noisy environment is still beyond the capabilities of current systems.

2. Augmentative communication systems

A perhaps surprising source of ideas for dealing with this is the field of augmentative and alternative communication (AAC) for people who are unable to speak through physical disability. Permanent inability to speak may result from cerebral palsy, degenerative conditions, or strokes. Approximately 0.1 % of the population has severe speech and communication problems.

An increasing range of computer-assisted systems is available for supporting augmentative and alternative communication [2]. Users can create text for speaking letter by letter, or word by word. This is very flexible, but has a considerable speed penalty. Because of their general physical disability, most users are restricted with these communication methods to rates of 2-10 words per minute.

A new generation of communication systems is emerging which rely on large amounts of prestored conversational material, organised in a way which reflects conversational structure and flow. It would be relatively easy to implement a multi-lingual capability with many of these systems, to the extent that they are based on a finite number of prestored language units. This would be particularly true of systems, typically research prototypes thus far, which are based on the rapid navigation through prestored conversational material by means of conversation structure and utterance prediction [3].

3. Modelling conversational routines

Finding ways to model conversational routines has been proposed as a means to enhance communication rate in AAC systems. A number of prototypes have been developed to illustrate this approach. These have involved, for example, providing the user with predicted texts following the normal conversational sequence of: opening, smalltalk, discussion, wrapping-up, and closing [4]. Story telling with a large number of pre-stored utterances has made it possible for users to express their personalities and achieve more satisfying conversations [5, 6, 7]. Focusing on transactional-type conversation is another recent approach, which is quite effective because of its provision of predictable communication sequences in everyday encounters [8, 9, 10].

As well as offering the potential to increase the communication rate of AAC users, the technique of prestored material along with conversation modelling could be extended to provide them with a multi-lingual capability. Although simpler than an automatic translation system, such an approach would have the advantage that, with material carefully translated beforehand, it would be possible to take into account language features such as idioms and slang, and also any cultural differences in language usage.

Such a system could just as easily be used by speaking people whose only communication disadvantage is not being able to speak another language. The potential cross-fertilisation of ideas from work on systems for people with disabilities and the wider field of human-computer interaction has been pointed out by Newell, who notes that systems developed to help a minority of people who need help in ordinary situations may also be of use for everyone, when they find themselves in an unusual situation or environment [11]. In this study, the AAC techniques described above were incorporated into a prototype which provide a multi-lingual capability for non-speaking people. The system also has the potential of wider use by people whose communication difficulty is simply that they do not speak another foreign language.

4. A prototype rapid multilingual communicator

The prototype held a large amount of pre-stored conversational material which was translated beforehand into several languages, taking full account of cultural as well as linguistic aspects of the utterances. It provided a set of specific scenarios, such as ‘at the airport’, ‘at a tourist information office’, ‘at a hotel’, and ‘at a hospital’. In addition it provided a symbol based communication facility for conveying simple messages by means of symbols. An on-screen keyboard was also available. The conversational material was arranged in accordance with anticipated frequency of use, conceptual grouping and conversation flow.

The prototype was designed to be operated by two speakers who have different languages. When the user is composing an utterance, the interface is displayed for them in their own language. When the user chooses to have the utterance spoken by the system, the utterance is spoken in the other language and then all the labels on the screen interface are translated into that language, ready for the other participant to make use of them. In addition to being multi-lingual, the interface is multimodal, combining text with symbols and pictures. It was hoped that this pictorially enhanced interface would help users to recognize the pre-stored utterances quickly and also make the system more accessible for non-literate people.

The system was developed in Java, which allows it to be run on different operating systems without major changes. It is also planned to eventually make use of the system in communicating over the internet.

The prototype is composed of two parts, the program and the data. Because the functions and interface of the program are all determined by the information in the data files, the stored conversation materials and the arrangement of the components can be modified simply by editing these data files without any alteration to the program.
The system is operated by both speakers who have different mother tongues or disabilities. While composing an utterance, this system is displayed in a speaker's own language, and when finished, the utterance and all the other components on the screen are automatically translated and spoken in the other language. The language pair can also be selected manually. This is sometimes necessary when a speaker wants to make a series of separate utterances before handing over the conversation turn to the other person, and thus needs to over-ride the automatic switching to the other language after an utterance. In addition to multilinguality, the output of the prototype has another feature: multimodality -- speech output with an interface which combines text, symbols and pictures. This multimodality could potentially provide a universal environment for all conversation participants regardless of the difference of communication method between the speakers.

The pragmatics of conversation determined the interface design of the prototype. Pre-stored conversation materials are arranged in accordance with the frequency of use, conceptual grouping and conversation flow. This assists the user in accomplishing the task quickly in a transactional-type conversation by following a familiar communication path.

5. Evaluation of the system

The system was evaluated by comparing it to its nearest equivalent: a multi-lingual phrase book which is available for Japanese tourists, and which has a large number of phrases in several languages, grouped by the setting in which they might be used.

This evaluation consisted of two phases, role plays and data analysis. In the first phase, twenty-four cross-language conversations were videotaped by four groups by using either the prototype or phrase book. Each group had three participants, English, Japanese, and Spanish native speakers. The languages were ordered at random. There were six scenarios based on frequently occurring conversations at an airport check-in counter. There were two roles in a conversation, the staff member and passenger, which were also performed in a random order.

In the second phase of the evaluation twenty native English speakers, who were different from the participants of the role plays, were asked to rate the conversations by filling out a questionnaire after reading the translated transcriptions of the recorded conversations.

6. Results

6.1. Speed of communication

In terms of the number of words conveyed per minute, it was concluded so far that there is no statistically significant difference between the two methods. This conclusion however results primarily from the large deviation of the data.

6.2. Volume of communication and frequency of initiative taking

The data from the evaluations showed that the participants using the prototype used it to produce conversational contributions more frequently than the participants using the phrase book. This phenomenon was marked particularly for the staff member.

6.3. Agreement and understanding

There was no statistically significant difference between the two methods in terms of the degree of agreement and understanding. The same result was obtained for the degree of success overall, although the prototype was better at accomplishing sub-tasks.

6.4. Social routines

In conversations with strangers, such as those that occur at check-in counters, in hotels, and in restaurants, politeness is helpful in accomplishing the tasks pleasantly and efficiently [12]. With respect to this point, it was found that the prototype was superior to the phrase book. The prototype might be improved in the future by displaying automatically phrases useful for social routines when necessary, since many of these routines must be used with correct timing to be effective.

7. Conclusions

The participants using this prototype produced more natural and successful conversations, particularly due to their ability to take the initiative and their increased use of social routines in the interactions. Thus it appears that computer-based systems such as the prototype have the potential to be more effective than printed material because of the possibilities inherent being able to model important features of a conversation and portray them easily in a dynamic display. It is notable that shared use of the same system by both speakers also conveys an advantage because the system can use the currently spoken utterance to produce/predict the subsequent response. The trials of this prototype have shown that if future communication aids for non-speaking people, regardless of their multi-lingual capability, could in some way be used equally by both partners in a dialogue, this might add considerably to their effectiveness.

This prototype provides an example of the potential for a wider application of technology which has been initially developed to assist people with disabilities. As well as providing a disabled non-speaking person with a multi-lingual capability, the same system may be also be used to provide an effective multi-lingual communication tool for speaking people.

References


