Converser™: Highly Interactive Speech-to-Speech Translation for Healthcare

Mike Dillinger
Spoken Translation, Inc.
Berkeley, CA, USA 94705
mike.dillinger@spokentranslation.com

Mark Seligman
Spoken Translation, Inc.
Berkeley, CA, USA 94705
mark.seligman@spokentranslation.com

Abstract

We describe a highly interactive system for bidirectional, broad-coverage spoken language communication in the healthcare area. The paper briefly reviews the system's interactive foundations, and then goes on to discuss in greater depth our Translation Shortcuts facility, which minimizes the need for interactive verification of sentences after they have been vetted. This facility also considerably speeds throughput while maintaining accuracy, and allows use by minimally literate patients for whom any mode of text entry might be difficult.

1 Introduction

Spoken Translation, Inc. (STI) of Berkeley, CA has developed a commercial system for interactive speech-to-speech machine translation designed for both high accuracy and broad linguistic and topical coverage. Planned use is in situations requiring both of these features, for example in helping Spanish-speaking patients to communicate with English-speaking doctors, nurses, and other healthcare staff.

The twin goals of accuracy and broad coverage have until now been in opposition: speech translation systems have gained tolerable accuracy only by sharply restricting both the range of topics which can be discussed and the sets of vocabulary and structures which can be used to discuss them. The essential problem is that both speech recognition and translation technologies are still quite error-prone. While the error rates may be tolerable when each technology is used separately, the errors combine and even compound when they are used together. The resulting translation output is generally below the threshold of usability – unless restriction to a very narrow domain supplies sufficient constraints to significantly lower the error rates of both components.

STI's approach has been to concentrate on interactive monitoring and correction of both technologies.

First, users can monitor and correct the speaker-dependent speech recognition system to ensure that the text, which will be passed to the machine translation component, is completely correct. Voice commands (e.g. **Scratch That** or **Correct <incorrect text>**) can be used to repair speech recognition errors. While these commands are similar in appearance to those of IBM's ViaVoice or ScanSoft's Dragon NaturallySpeaking dictation systems, they are unique in that they will remain usable even when speech recognition operates at a server. Thus, they will provide for the first time the capability to interactively confirm or correct wide-ranging text, which is dictated from anywhere.

Next, during the MT stage, users can monitor, and if necessary correct, one especially important aspect of the translation – lexical disambiguation.

STI's approach to lexical disambiguation is twofold: first, we supply a specially controlled *back translation*, or translation of the translation. Using this paraphrase of the initial input, even a monolingual user can make an initial judgment concerning the quality of the preliminary machine translation output. To make this technique effective, we use proprietary facilities to ensure that the

Proceedings of the Workshop on Medical Speech Translation at HLT-NAACL 2006, pages 40–43,
New York, New York, June 2006.

lexical senses used during back translation are appropriate.

In addition, in case uncertainty remains about the correctness of a given word sense, we supply a proprietary set of Meaning Cues™ – synonyms, definitions, etc. – which have been drawn from various resources, collated in a unique database (called SELECT™), and aligned with the respective lexica of the relevant machine translation systems. With these cues as guides, the user can select the preferred meaning from among those available. Automatic updates of translation and back translation then follow.

The result is an utterance, which has been monitored and perhaps repaired by the user at two levels – those of speech recognition and translation. By employing these interactive techniques while integrating state-of-the-art dictation and machine translation programs – we work with Dragon Naturally Speaking for speech recognition; with Word Magic MT (for the current Spanish system); and with ScanSoft for text-to-speech – we have been able to build the first commercial-grade speech-to-speech translation system which can achieve broad coverage without sacrificing accuracy.

2 Translation Shortcuts

In order to accumulate translations that have been verified by hand and to simplify interaction with the system, we have developed additional functionality called Translation Shortcuts™.

Shortcuts are designed to provide two main advantages:

First, re-verification of a given utterance is unnecessary. That is, once the translation of an utterance has been verified interactively, it can be saved for later reuse, simply by activating a **Save as Shortcut** button on the translation verification screen. The button gives access to a dialogue in which a convenient *Shortcut Category* for the Shortcut can be selected or created. At reuse time, no further verification will be required. (In addition to such dynamically created *Personal* Shortcuts, any number of prepackaged *Shared* Shortcuts can be included in the system.)

Second, access to stored Shortcuts is very quick, with little or no need for text entry. Several facilities contribute to meeting this design criterion.

- A *Shortcut Search* facility can retrieve a set of relevant Shortcuts given only keywords or the first few characters or words of a string. The desired Shortcut can then be executed with a single gesture (mouse click or stylus tap) or voice command.

NOTE: If no Shortcut is found, the system automatically gives access to the full power of broad-coverage, interactive speech translation. Thus, a seamless transition is provided between Shortcuts and full translation.

- A *Translation Shortcuts Browser* is provided, so that users can find needed Shortcuts by traversing a tree of Shortcut categories. Using this interface, users can execute Shortcuts even if their ability to input text is quite limited, e.g. by tapping or clicking alone.

The demonstration will show the Shortcut Search and Shortcuts Browser facilities in use. Points to notice:

- The Translation Shortcuts Panel contains the Translation Shortcuts Browser, split into two main areas, Shortcuts Categories (above) and Shortcuts List (below).
- The Categories section of the Panel shows the current selected category, for example **Conversation**, which contains everyday expressions. This category has a **Staff** subcategory, containing expressions most likely to be used by healthcare staff members. There is also a **Patients** subcategory, used for patient responses. Such categories as **Administrative topics** and **Patient's Current Condition** are also available; and new ones can be freely created.
- Below the Categories section is the Shortcuts List section, containing a scrollable list of alphabetized Shortcuts. (Various other sorting criteria will be available in the future, e.g. sorting by frequency of use, recency, etc.)
- Double clicking on any visible Shortcut in the List will execute it. Clicking once will select and highlight a Shortcut. Typing **Enter** will execute the currently highlighted Shortcut, if any.
- It is possible to automatically relate options for a patient's response to the previous staff member's utterance, e.g. by automatically going to the sibling **Patient** subcategory if the prompt was given from the **Staff** subcategory.

Because the Shortcuts Browser can be used without text entry, simply by pointing and clicking, it enables responses by minimally literate users. In the future, we plan to enable use even by completely illiterate users, through two devices: we will enable automatic pronunciation of Shortcuts and categories in the Shortcuts Browser via text-to-speech, so that these elements can in effect be read aloud to illiterate users; and we will augment Shared Shortcuts with pictorial symbols, as clues to their meaning.

A final point concerning the Shortcuts Browser: it can be operated entirely by voice commands, although this mode is more likely to be useful to staff members than to patients.

We turn our attention now to the Input Window, which does double duty for Shortcut Search and arbitrary text entry for full translation. We will consider the search facility first.

- Shortcuts Search begins automatically as soon as text is entered by any means – voice, handwriting, touch screen, or standard keyboard – into the Input Window.
- The **Shortcuts Drop-down Menu** appears just below the Input Window, as soon as there are results to be shown. The user can enter a few words at a time, and the drop-down menu will perform keyword-based searches and present the changing results dynamically.
- The results are sorted alphabetically. Various other sorting possibilities may be useful: by frequency of use, proportion of matched words, etc.
- The highest priority Shortcut according to the specified sorting procedure can be highlighted for instant execution.
- Highlighting in the drop-down menu is synchronized with that of the Shortcuts list in the Shortcuts Panel.
- Arrow keys or voice commands can be used to navigate the drop-down menu.
- If the user goes on to enter the exact text of any Shortcut, e.g. "Good morning," a message will show that this is in fact a Shortcut, so that verification will not be necessary. However, final text not matching a Shortcut, e.g. "Good job," will be passed to the routines for full translation with verification.

3 Future developments

We have already mentioned plans to augment the Translation Shortcuts facility with text-to-speech and iconic pictures, thus moving closer to a system suitable for communication with completely illiterate or incapacitated patients.

Additional future directions follow.

- **Server-based architectures:** We plan to move toward completely or partially server-based arrangements, in which only a very thin client software application – for example, a web interface – will run on the client device. Such architectures will permit delivery of our system on smart phones in the Blackberry or Treo class. Delivery on handhelds will considerably diminish the issues of physical awkwardness discussed above, and anytime/anywhere/any-device access to the system will considerably enlarge its range of uses.
- **Pooling Translation Shortcuts:** As explained above, the current system now supports both Personal (do-it-yourself) and Shared (prepackaged) Translation Shortcuts. As yet, however, there are no facilities to facilitate pooling of Personal Shortcuts among users, e.g. those in a working group. In the future, we will add facilities for exporting and importing shortcuts.
- **Translation memory:** Translation Shortcuts can be seen as a variant of Translation Memory, a facility that remembers past successful translations so as to circumvent error-prone reprocessing. However, at present, we save Shortcuts only when explicitly ordered. If all other successful translations were saved, there would soon be far too many to navigate effectively in the Translation Shortcuts Browser. In the future, however, we could in fact record these translations in the background, so that there would be no need to re-verify new input that matched against them. Messages would advise the user that verification was being bypassed in case of a match.
- **Additional languages:** The full SLT system described here is presently operational only for bidirectional translation between English and Spanish. We expect to expand the system to Mandarin Chinese next. Limited working prototypes now exist for Japanese and German, though we expect these languages to be most useful in application fields other than healthcare.

4 Conclusion

We have described a highly interactive system for bidirectional, broad-coverage spoken language communication in the healthcare area. The paper has briefly reviewed the system's interactive foundations, and then gone on to discuss in greater depth issues of practical usability.

We have presented our Translation Shortcuts facility, which minimizes the need for interactive verification of sentences after they have been vetted once, considerably speeds throughput while maintaining accuracy, and allows use by minimally literate patients for whom any mode of text entry might be difficult.

We have also discussed facilities for multimodal input, in which handwriting, touch screen, and keyboard interfaces are offered as alternatives to speech input when appropriate. In order to deal with issues related to physical awkwardness, we have briefly mentioned facilities for hands-free or eyes-free operation of the system.

Finally, we have pointed toward several directions for future improvement of the system.