Automatic Speech Translation for Healthcare: Some Internet and Interface Aspects

Mark Seligman
Spoken Translation, Inc.
1100 West View Drive
Berkeley, CA 94705
mark.seligman@spokentranslation.com

Mike Dillinger
Spoken Translation, Inc.
1100 West View Drive
Berkeley, CA 94705
mike.dillinger@spokentranslation.com

Abstract

We describe Converser for Healthcare, Version 4.0, a real-time, multi-modal, broad-coverage, highly interactive translation system. Version 3.0 was successfully tested in three departments of a large hospital complex belonging to a major US healthcare organization. Based on lessons learned, some implications of online use are discussed, along with selected interface issues.

1 Introduction

Demand for interpretation in healthcare settings between English and other languages has been increasing in recent years. San Francisco General Hospital, for example, receives more than 42,000 requests per year for 35 different languages, distributed among many clinics (Paras, et al., 2002).

In response, several groups have experimented with systems for automatic spoken language translation (Bouillon, Ehsani et al., 2006, 2008).

This paper discusses aspects of a real-time, multi-modal, broad-coverage, highly interactive translation system, Converser for Healthcare, Version 4.0. Version 3.0 has been tested in three departments of a large hospital complex belonging to a major US healthcare organization (Seligman and Dillinger, 2011).

Section 2 will briefly describe the Converser system, sketching its approach to highly interactive real-time translation. For fuller description, see (Dillinger and Seligman, 2004); (Zong and Seligman, 2005); and (Seligman and Dillinger 2006, 2008, 2011).

Based upon lessons learned, Section 3 will discuss some aspects of online use in the current version; and Section 4 will address selected interface issues in this version. We conclude in Section 5.

2 The Converser System

We now briefly summarize our approach to real-time automatic interpretation.

Speech-enabled translation systems face an essential problem: both speech recognition and translation technologies are still quite error-prone. The errors combine and even compound when they are used together, so that the resulting translation output may be unusable unless restriction to a narrow domain supplies sufficient constraints.

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

First, users can monitor and correct the speech recognition system to ensure that the text which will be passed to the machine translation component is completely correct.

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

The system's approach to lexical disambiguation is twofold: first, we supply a *Back-Translation*, or re-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. Other systems, e.g. IBM's MASTOR (Gao, Liang, et al., 2006), have also employed re-translation. Converser, however, exploits proprietary technologies to ensure that the lexical senses used during back translation accurately reflect those used in forward translation.

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

Such interactivity within a speech translation system can provide increased accuracy and confidence, even for wide-ranging conversations (Zong and Seligman, 2005).

Translation Shortcuts. The Converser system includes Translation Shortcuts™ – pre-packaged translations, designed to provide two main advantages:

First, re-verification of a given utterance is unnecessary, since it has been pre-translated by a professional (or, in future versions of the system, verified using the system's feedback and correction tools).

Second, access to stored Shortcuts is very quick, with little or no need for text entry. Two facilities contribute to quick access:

Shortcut Search 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 or voice command. If no Shortcut is found to match the input text, the system automatically and seamlessly gives access to broad-coverage, interactive speech 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 by tapping or clicking.

Figure 1 shows the **Shortcut Search** and **Shortcuts Browser** facilities in use.

- On the left, the **Translation Shortcuts Panel** contains the **Translation Shortcuts Browser**, split into two main areas, **Shortcuts Categories** (above) and **Shortcuts List** (below).
- The Categories section shows current selection of the **Conversation** category, containing everyday expressions. Categories for **Administrative topics** and **Patient's Current Condition** are also visible.
- The **Shortcuts List** contains a scrollable list of alphabetized Shortcuts in the selected Category.

The **Input Window** does double duty for **Shortcut Search** and entry of text for full translation. The search facility is shown in Figure 2.

- **Shortcuts Search** begins automatically as soon as text is entered – by voice, handwriting, touch screen, or standard keyboard.
- The **Shortcuts Drop-down Menu** appears just below the Input Window. Here, the user has entered "Good" and a space, so the search program has received its first input word. The drop-down menu shows the results of a keyword-based search, with the first hit highlighted for instant execution.
- If the user goes on to enter the exact text of any Shortcut, e.g. "Good morning," the interface will confirm Shortcut recognition. Final text not matching a Shortcut, e.g. "Good job," will undergo full translation with verification.

Multimodal input. In addition to dictated speech, we enable handwritten input, the use of touch screen keyboards for text input, and the use of standard keyboards.

Having sketched the Converser speech translation system, we go on to discuss aspects of its online use and interface.

3 Internet Issues

Converser's current Internet aspects can be considered from two viewpoints – real-time and offline.

Real-time Internet issues relate to the practical trade-offs between device-based and cloud-based services. The system is in transition from the former to the latter, and is now in a hybrid condition, with most of its software running on the desktop, but its speech recognition element running in the cloud.

Why cloud-based ASR? To enable speaker-independence for very large ASR vocabularies. Large vocabularies are required because of the system's emphasis on wide-ranging coverage. Until now, to enable this breadth, each user has had to create an individual voice profile – a process requiring some training and in recent years taking between three and ten minutes, depending on the degree of initial personalization. While the associated delay for healthcare staff members has remained tolerable for small-scale trials, scaling to widespread use would be problematic. Worse, however, has been the impracticality of organizing such enrollment for patients, especially those using the system once only, e.g. at a pharmacy window. Thus Spanish speakers have had only occasional access to Converser's speech input – a major issue in Version 3.0.

Since about 2010, however, dictation systems have become commercially available which include acoustic models built from very diverse training samples, thus giving good results without the need for initial enrollment. A solution for Converser's most pressing issue seemed to be at hand.

But a final snag appeared. Current user-independent dictation components have been developed for *cloud-based* use; but Converser 4.0 remains a *desktop* system for now. Accordingly, we employ an adapter offered by a third party (SpeechTrans, Inc.) which enables cloud-based speech recognition for desktop use. Given a good Internet connection, robust speaker-independent results are achieved for both English and Spanish; and a major system issue is finally resolved.

Nothing is perfect, however. As a partly cloud-based system, Converser now requires internet connectivity; but this is not guaranteed for every customer organization or use case. The host of Converser's pilot project, for example, is only now moving toward a company-wide WiFi standard. As this move is completed, a wholly cloud-based Converser version will be created, to be used as an especially mobile alternative to the desktop model, and ultimately replacing it.

Offline Internet issues relate to the resources which support Converser's interactive translation facilities.

As explained above, to enable choice of word sense for ambiguous words, the system supplies Meaning Cues™ – synonyms, definitions, etc. – drawn from various resources, collated in a database (called SELECT™), and aligned with the lexica of the current MT system.

The cues are collected from various online sources – WordNet and others – and sorted by meaning, using proprietary algorithms. The result is a database of cue sets which must then be aligned with the word meanings (acceptions) of the current translation system, again via proprietary methods. Groupings and alignments are then checked by linguists.

Thus far, Converser has incorporated only rule-based translation components, so sorting and alignment methods have concentrated on these. Recent work has focused on extending these methods to statistical MT engines.

4 Interface Issues

Based on the results of our pilot test, a previous paper (Seligman and Dillinger, 2011) noted the need for several software and interface improvements,. We take this opportunity to report on selected upgrades and refinements.

- **Speaker-independent ASR:** The most important improvement has already been discussed: speech recognition is now available without enrollment for both input languages.
- **On-screen mic:** In support of the new ASR, an on-screen mic button has been added, thus eliminating the onerous task of setting up tablet buttons for each new speech recognition user. However, users and use cases may differ in their preferences for using the new mic button. It may sometimes be preferable to click once to turn the mic on, and click again to turn it off; for other tasks, the preference may be a push-to-talk functionality; for yet others, click-on-but-automatic-off may be most convenient. All of these mic modes have now been implemented. In the future, an auto-on-auto-off mode will be added to enable entirely hands-and-eyes-free use of the system.
- **Translation verification modes:** In pilot trials, while Converser's verification features proved important for accuracy and confidence in serious use cases, they also slowed the interchange, often unnecessarily in less serious use cases. We have thus enabled an optional mode in which translation proceeds without a

pause for verification. A Traffic Light icon is used in this case to signal "Full speed ahead!" via a green light. (Back-Translation is now included in the on-screen and saveable transcript, however, so that translation accuracy can still be checked *after* the transmission to enable follow-up clarification. This post-check remains even when the green light is on.) Switching the light to yellow signals "Proceed with caution!", and returns to the slower but safer mode. The red light setting blocks all translation.

- **ASR verification modes:** The Traffic Light controls verification of *translation* per se; but verification of *speech recognition* is a separate matter, and can be separately controlled. In ASR Green Light Mode, translation automatically begins when the microphone is switched off. In ASR Yellow Light Mode, the system waits for an explicit signal to begin translation, thus giving an opportunity to correct ASR before proceeding.
- **Text-to-speech:** Users can now control the speed of text-to-speech, especially to slow it when desired.
- **Introducing Converser:** The *How to Use Converser* Translation Shortcut category has been refined to provide a smoother introduction for first-time users. An instructional video is planned.

5 Conclusions

We have discussed aspects of Converser for Healthcare 4.0, a real-time, multi-modal, broad-coverage, highly interactive translation system. Emphasis has been upon aspects of online use in the current version, specifically on (1) the need for cloud-based ASR and (2) the online collection of cues such as synonyms and definitions to support the system's interactive translation facilities. We have also examined selected interface issues in the new version, describing improvements made in response to pilot tests.

Acknowledgments

The authors thank SpeechTrans, Inc. for software support as described herein.

References

Pierrette Bouillon, Farzad Ehsani, et al. 2006. *Proceedings of the First International Workshop on Medical Speech Translation*, New York, NY: Association for Computational Linguistics, June, 2006..

Pierrette Bouillon, Farzad Ehsani, et al. 2008. *COLING 2008: Proceedings of the workshop on Speech Processing for Safety Critical Translation and Pervasive Applications*, COLING 2008, Manchester, UK, August, 2008.

Mike Dillinger and Mark Seligman. 2004. A highly interactive speech-to-speech translation system. In *Proceedings of the VI Conference of the Association of Machine Translation in the Americas*. E. Stroudsburg, PA: American Association for Machine Translation.

Yuqing Gao, Gu Liang, Bowen Zhou, Ruhi Sarikaya, Mohamed Afify, Hong-Kwang Kuo, Wei-zhong Zhu, Yonggang Deng, Charles Prosser, Wei Zhang and Laurent Besacier. (2006). IBM MASTOR system: multilingual automatic speech-to-speech translator. In: *HLT-NAACL 2006: Proceedings of the Workshop on Medical Speech Translation* . New York, NY, USA.

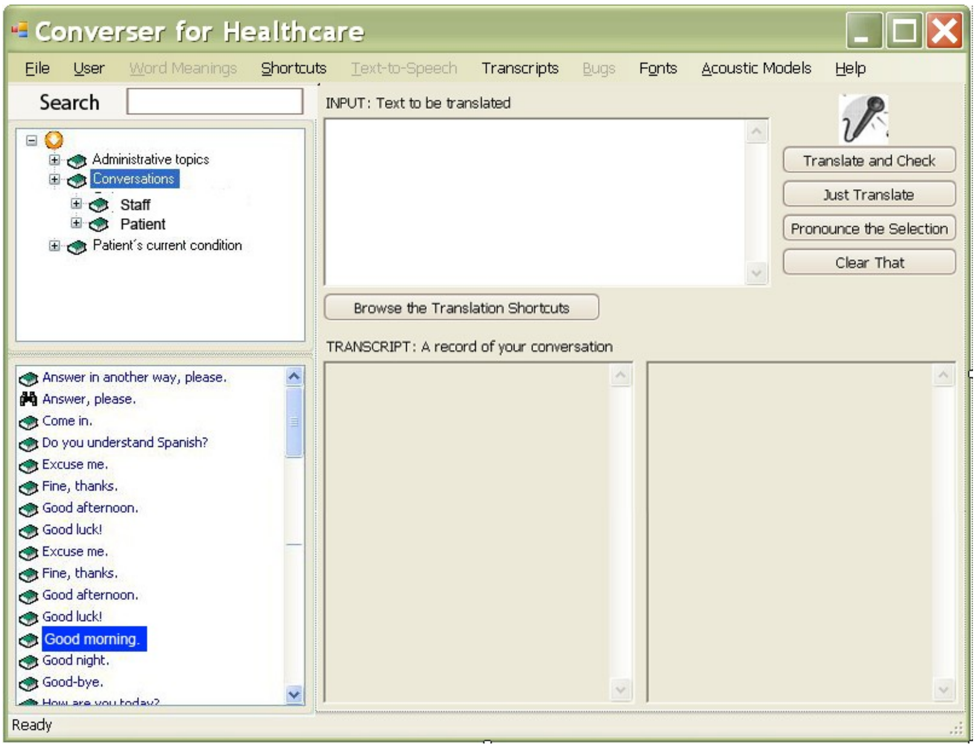
Melinda Paras, O. Leyva, T. Berthold, and R. Otake. 2002. *Videoconferencing Medical Interpretation: The results of clinical tri*als. Oakland, CA: Heath Access Foundation.

Mark Seligman and Mike Dillinger. 2006. Usability issues in an interactive speech-to-speech translation system for healthcare. In *Proceedings of the First International Workshop on Medical Speech Translation.*. New York, New York: Association for Computational Linguistics, June, 2006.

Mark Seligman and Mike Dillinger. 2008. Rapid portability among domains in an interactive spoken language translation systgem. *COLING 2008: Proceedings of the workshop on Speech Processing for Safety Critical Translation and Pervasive Applications*. Manchester, UK: August, 2008.

Mark Seligman and Mike Dillinger. 2011. Real-time Multi-media Translation for Healthcare: a Usability Study. In *Proceedings of MT Summit 2011: Machine Translation Summit XIII*. Xiamen, China: September, 2011.

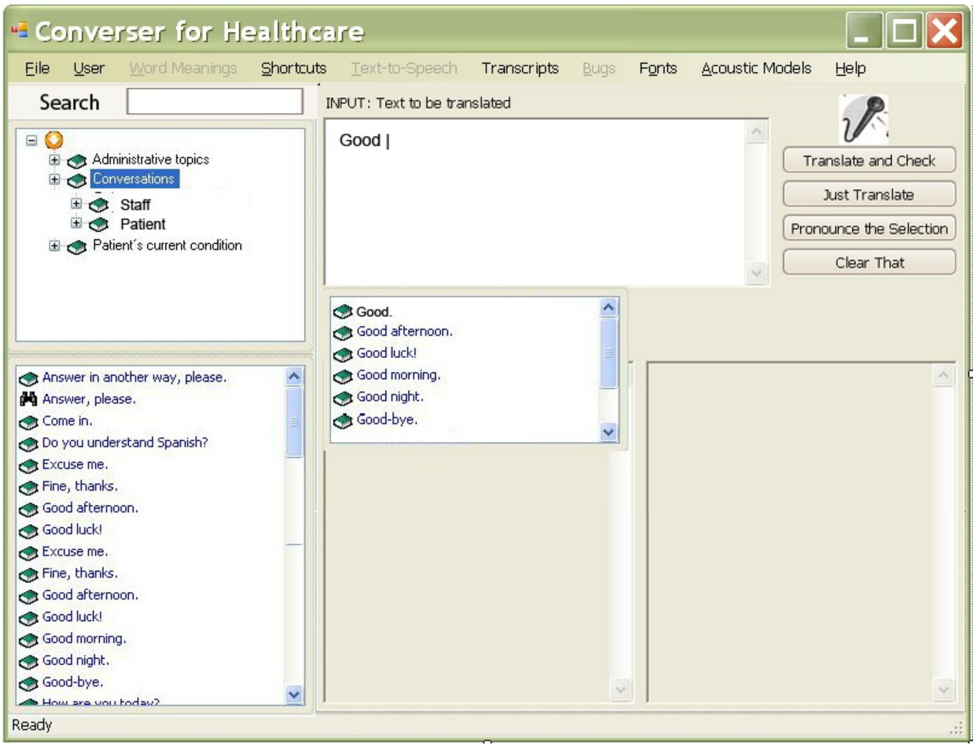
Chengqing Zong and Mark Seligman. 2005. Toward Practical Spoken Language Translation. *Machine Translation,* 19(2): 113-137. June, 2005.



Administrative topics
Conversations
Staff
Patient
Patient's current condition

Answer in another way, please.
Answer, please.
Come in.
Do you understand Spanish?
Excuse me.
Fine, thanks.
Good afternoon.
Good luck!
Excuse me.
Fine, thanks.
Good afternoon.
Good luck!
Good morning.
Good night.
Good-bye.
How are you today?

Figure 1: The Input Screen, showing the Translation Shortcuts Browser and Search facilities.



Administrative topics
Conversations
Staff
Patient
Patient's current condition

Good |

Good.
Good afternoon.
Good luck!
Good morning.
Good night.
Good-bye.

Answer in another way, please.
Answer, please.
Come in.
Do you understand Spanish?
Excuse me.
Fine, thanks.
Good afternoon.
Good luck!
Excuse me.
Fine, thanks.
Good afternoon.
Good luck!
Good morning.
Good night.
Good-bye.
How are you today?

Figure 2: The Input Screen, showing automatic keyword search of the Translation Shortcuts.