Spoken Language Translation: Three Business Opportunities

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

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

Abstract

This paper reports on three business opportunities encountered by Spoken Translation, Inc., a developer of software systems for automatic spoken translation: (1) a healthcare organization needing improved communications between limited-English patients and their caregivers; (2) a networking and communications firm aiming to add UN-style simultaneous interpreting to their telepresence facilities; and (3) the retail arm of a device manufacturer hoping to enable more effective in-store consulting for customers with imperfect command of an outlet's native language. None of these openings has yet led to substantial business, but one remains in negotiation. We describe how the business introductions came to us; the proposed use cases; demonstrations, presentations, tests, etc.; and issues/challenges. We also comment on early consumer-oriented products for spoken language translation. The aim is to provide a snapshot of one company's business possibilities and challenges at the dawn of the era of automatic interpreting.

1 Introduction

Spoken language translation (SLT) or automatic interpreting is still a very new and immature technology. While it's clear that the demand for human-level SLT would be immense in the face of relentless globalization, it's still challenging to match specific current business use cases with state-of-the-art capabilities. This paper reports on three business opportunities encountered by Spoken Translation, Inc. (STI), a developer of SLT software systems. We won't identify the prospects, but can describe them generically as (1) a healthcare organization needing improved communications between limited-English patients and their caregivers; (2) a networking and communications firm aiming to add UN-style simultaneous interpreting to their telepresence facilities; and (3) the retail arm of a device manufacturer hoping to enable more effective in-store consulting for customers with imperfect command of an outlet's native language. None of these openings has yet led to substantial business; but one remains in negotiation, and the others can be revisited as the technology and our company's capabilities mature. For each opportunity, we describe how the business introductions came to us; the proposed use case; what sorts of demonstrations, presentations, and tests were requested and delivered; and our impressions concerning issues/challenges to be faced in order to close these or comparable deals in the future. We'll also comment on the relation between these use cases and the early consumer-oriented products now ascendant in the nascent SLT commercial field. Overall, the aim will be to provide a snapshot of one company's business possibilities and challenges at the dawn of the era of automatic interpreting.

Section 2 of this paper will review Converser, STI's real-time automatic translation system. Sections 3, 4, and 5 will report on the healthcare, telepresence (business-to-business), and retail (business-to-customer) opportunities respectively. We discuss and conclude in a final section.

2 The Converser System

We now briefly summarize STI's approach to real-time automatic interpretation in its Converser system.

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

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. Voice commands (e.g. **Scratch That** or **Correct <incorrect text>**) can be used to repair speech recognition errors.

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 (Seligman & Dillinger, 2004).

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 (mouse click or stylus tap) 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.

The Input Window does double duty for Shortcut Search (by initial characters or by keywords) and for entry of text for full translation.

Multimodal input. Speech input isn't appropriate for every situation, so Converser provides several input modes. In addition to dictated speech, we enable handwritten input, the use of touchscreen keyboards for text input, and the use of standard keyboards. All of these input modes are completely bilingual, and language switching is arranged automatically when there is a change of active participant. Further, it is possible to change input modes seamlessly within a given utterance: for example, users can dictate the input if they wish, but then can make corrections using handwriting or one of the remaining two modes.

Having surveyed the Converser system, we now go on to discuss three of its business opportunities.

3 Healthcare

In March, 2008, an investment firm, seeking validation of the demand for Converser, referred STI to a Vice President of Innovation and Advanced Technology at a large healthcare organization. She in turn passed us to a physician serving as Senior Technology Analyst, Innovation and Advanced Technology; we met in November of that year. And he in turn passed us to the Director, National Linguistics and Cultural Programs for a meeting in February – almost a year after our introduction to the organization. Interest was shown at each stage, so the decision was made to introduce staff members closer to the line of fire.

With the mediation of the Director of Linguistics and Cultural Services for a large urban area, we made several presentations over the next year and a half to groups of managers concerned with cultural issues. In November, 2009, STI was the only outside vendor invited to host a booth at the organization's National Diversity Conference. The receptions at all of these presentations were warm – at the Diversity Conference, more than eighty attendees requested additional information – so the high-level staff sought ways to fund more formal next steps. Through the advocacy of the Technology Analyst, we were introduced to the Director of Information Technology, Division of Research, and subsequently invited to formally propose a pilot project to the organization's Innovation Fund. We did, with close cooperation from the urban area's Director. Approval was received in January, 2011 – by then, almost three years after our introduction.

The pilot project ran for nine calendar months in 2011, with Converser use in three departments – pharmacy, in-patient nursing, and eye care – during three of those months. The six-person project team, including the Technology Analyst, representatives of the Innovation Fund, and technical specialists, met weekly over several months as well. See (Seligman and Dillinger, 2011) for a fuller account of the pilot project, including discussion of issues and lessons learned. Also during the project, STI was invited to speak to, and demo for, a group of some thirty interested parties at the organization's showcase and research center; and center staff independently demonstrated Converser in several locations.

The pilot concluded with sixty-one interviews with patients and staff members who used Converser, carried out by an interpreter from an outside agency. A formal internal report gave the results. When asked, “Did [Converser] meet your needs?” 94% of the respondents answered either Completely or Mostly, and 90% judged translation accuracy to be High.

These pilot project results were presented to a small group by the Innovation Fund liaison, whose work was then concluded. There followed a lull, partly occasioned by an illness and a dismissal (for extraneous reasons) of the two principal advocates. However, following a meeting with the Director, National Linguistics and Cultural Programs and two colleagues, STI has once again been invited to play a featured role in a National Diversity Conference, which this year will present a hands-on exposition of coming technologies related to linguistic and cultural competence. While emphasizing facilities for reliability and customization, STI anticipates cooperation with one or more SLT app makers in order to demonstrate mobile speech-enabled translation on smartphones and new generation tablets. The Conference will convene in November, 2012, some four and a half years after STI's introduction to the healthcare organization.

4 Telepresence (Business-to-business)

In 2009, STI received a cold call from a large networking and communications company, referred by the vendor which supplies our speech recognition and text-to-speech. The networking company provides a telepresence product for corporate use – a kind of advanced multiparty Skype for corporations. Users sit around a semicircular table in a special-purpose studio. A large high-definition video screen abuts the table and displays a complementary half-table in a distant location, so that all participants appear to be sharing a single round table. Each participant uses a dedicated table microphone. STI's mission: to add free-flowing simultaneous speech to speech translation in the style of an international conference with human interpreters. (Continuous sub-titles were also desired.)

In view of the state of the art, as seen for example in project GALE (Cohen, 2007) or in the simultaneous speech translation system by Waibel and collaborators (Waibel, 2012), it was clear that perfect performance could hardly be expected. However, several factors were in the project's favor. The sound quality would be ideal under such studio conditions; the participants would be professionals comfortable with technology and usually speaking a standard dialect; and the areas of business discussion would be relatively predictable, giving an opportunity for tuning.

But we also had another trick up our sleeves: given STI's interactive approach, we were in a position to offer a combination of transparent or simultaneous and interactive speech translation. Users could (we proposed) freely ignore the automatic interpreting system, accepting a relatively high error rate (we predicted between 20% and 45%); however, when necessary for clarification, they could press a button to “proceed with caution” – to interrupt the translation flow in order to verify and when necessary correct, thus obtaining translation accuracy commensurate with the interaction time spent.

Three noteworthy demands were made of us during the evaluation process (along with many extensive presentations and discussions and a due diligence questionnaire).

First, we were asked to translate a previously recorded sound file from English into Spanish. We used Dragon NaturallySpeaking to produce English text, and then ran the text through our rule-based MT without user interaction, tuning over the course of a week using a glossary tool.

Second, we were asked to demo Converser in the telepresence environment. Technically, this proved to be simple, as the text-to-speech output from the Converser tablet computer could be plugged into a general-purpose audio jack in the telepresence system. A simulated English<> Spanish conversation was arranged: the first author spoke Spanish with interactive correction as necessary, with immediate translation into spoken English. Corporate staff members spoke English from a separate location. (Since they spoke without translation, the demo was in this sense one-sided.) The conversation, with perhaps a dozen inputs on each side, was recorded for later examination. We responded to improvised questions on general business and everyday social topics. Our interlocutors judged the responses to be acceptable in every case, and generally seemed enthusiastic: grinning, they volunteered that we were clearly adding value to the system.

Third and finally, we sketched a design for free-flowing speech translation and illustrated it with an animated slide. The incoming speech signal was to be segmented by pauses of a predefined length, and then successively queued for speech recognition, translation, and pronunciation/text display. (No explicit provision was made for interactive interruptions at this stage, however.)

Our immediate contacts appeared satisfied with these steps, and introduced us to their supervisor and her team. She praised our demo, then raised a few more questions by e-mail. We proposed a project of a few months to implement an API according to the company's specifications and to work closely with the company's engineers to arrive at a working system.

And that was the end of the line. Only very brief e-mails reached us thereafter, and we were left to intuit the reasons that no further progress was made. We guess that we were judged not quite ready for prime time – because our API was incomplete and because we were pre-revenue. There were also hints of business issues with the intended speech recognition vendor.

5 Retail (Business-to-customer)

The second author consults for a device manufacturer with retail outlets in several countries. In late 2011, his contact, a marketing executive, reached out to STI. She wanted to enable more effective in-store consulting for customers with imperfect command of a store's native language, hoping that a suitable app would also help present the company as internationally oriented. The relevant programs should be handy and require little or no training.

For maximum handiness and brand appropriateness, the programs should run on the company's native smartphones or tablets. This requirement was a first obstacle, since Converser runs to date on Windows-based machines; however, our contact was willing to entertain initial demos on this platform with a view toward porting later.

Another aspect of handiness played to our strengths, however. Consultants must often respond repeatedly to the same questions or problems, so our proprietary Translation Shortcuts^{TM} facility would be advantageous. As explained earlier, shortcuts are pre-translated phrases, arranged in categories in the manner of a phrasebook, which can be browsed or searched for instant execution. New shortcuts and categories can be quickly created to customize Converser applications for particular customers or markets. Shortcuts are integrated into the system as translation memory: if an input is recognized as a Shortcut, its prepared translation is presented immediately; otherwise, full translation is invoked. See (Seligman and Dillinger, 2006) for extensive discussion in the healthcare context.

Based upon a series of interviews with retail staff, we developed several new Shortcut categories. Incorporating these, we designed a demo script on request.

Our contact arranged a demo for her supervisor, with the first author taking the part of in-store consultant and herself as Spanish-speaking customer. It went smoothly, and once again we proposed a project of a few months, in which a trial roll-out would be made to several stores. Once again, however, we found ourselves blocked at the supervisory level; and once again we were left to guess the reason. Best guesses: inability to demonstrate on a the company's platform and inability to offer a turnkey solution.

Frustratingly, we were contacted again some six months later by the same marketing executive, inquiring about our progress. We could report no progress on the relevant platform, however: lacking capital to develop on spec, we were obliged to propose development on demand.

6 Discussion and Conclusions

When STI began operations in April, 2002, speech-enabled automatic translation was already the subject of considerable research, but no attempts had yet been made to commercialize the technology. The scientific and technical problems in making a system work beyond very narrow domains, even in the laboratory, were challenging enough: of the three major components – ASR, MT, and TTS, only the last was really ready for industrial use. However, the purely *business* problems were also formidable, centering on issues of distribution and demand.

Distribution: How would products reach users/customers? On what platforms? Through which sales channels?

Demand: What would be the initial use cases, given the inevitable technical and practical limitations of early products? As always, the search would be for potential users with a need *and* the ability to pay.

The subsequent ten years have seen dramatic growth in the readiness of ASR and MT. Equally important for the SLT business, however, they have also witnessed a breathtaking growth in the possibilities for distribution. The infrastructure is to support mass use of SLT – mobile computing and communications; instant messaging and texting; social networking; smartphones; telepresence; and VoIP – is now fully in place. Building on these elements, thousands of apps are now on sale – including several SLT apps (by Google, SpeechTrans, Vocre, and SayHi!). Overall, the questions regarding platforms and sales channels for SLT are much less problematic ten years on.

What about the demand? All of the current SLT apps are consumer-oriented; but the experiences recounted here demonstrate a clear demand for SLT in serious and monetizable use cases as well – in healthcare, in business-to-business communications, in business-to-customer retail, and plausibly by extension in many other vertical markets.

So why aren't serious SLT systems for vertical markets already in widespread use? From a technical viewpoint, we believe that reliability and customization facilities like STI's are necessary to offer customers sufficient confidence and convenience to adopt several new technologies – to cross several chasms, in marketspeak. While STI's experience shows that such facilities can indeed be provided, issues of scaling remain: provision can be made for English<>Spanish, but for the dozens of languages and hundreds of translation paths now offered by Google?

More relevant to the three opportunities discussed here are problems of organizational inertia on one hand and, on the other, problems of readiness related to capitalization.

With respect to inertia, the length of the healthcare sales cycle is evident in the ongoing story of our healthcare opportunity. On the bright side, the organization in question is now moving – deliberately but clearly – toward mobile and cloud-based IT. This trend will make its use of SLT technologies much more practical, and may at the same time offer a path toward more grass-roots or viral adoption, partially alleviating the need for top-down approval at every step.

With respect to readiness and capitalization, we appear to have missed our first shots at the telepresence and retail markets at least partly because we had no turn-key solutions to sell. We didn't have them because we lacked the necessary capital to prepare them in advance on spec; and we lacked the capital in part because investors worried that we lacked large customers – a classic double bind.

Overall, a more agile approach seems called for. While continuing our pursuit of large customers able to afford full-service SLT products, it would seem advisable to develop a revenue base with entry-level products in the app market, so that the resulting revenue can be invested in vertical market development.

In any case, with issues of distribution now largely resolved and with demand now clearly demonstrated, STI believes that SLT systems for serious use cases in vertical markets will soon appear alongside the nascent crop of consumer-oriented SLT apps.

Acknowledgments

Confidentiality agreements currently prevent disclosure of the identity of the companies discussed here. The authors look forward to expressing their gratitude to the participants in future publications.

References

Jordan Cohen. 2007. “The GALE project: A description and an update.” In IEEE Workshop on Automatic Speech Recognition and Understanding. ASRU. December 9-13, 2007.

Mike Dillinger and Mark Seligman. 2004. “A highly interactive speech-to-speech translation system.” In *Proceedings of the VI Conference of the Association for Machine Translation in the Americas*. Washington, D.C., September-October, 2004.

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, June, 2006.

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,* Association for Computational Linguistics. New York, NY, June, 2006.

Mark Seligman and Mike Dillinger. 2011. “Real-time Multi-media translation for healthcare: a usability study.” In *Proceedings of the 13th Machine Translation Summit,* Xiamen, China, September 19-23, 2011.

Alexander Waibel. 2012. http://innovation.mfg.de/en/news-and-features/simultaneous-translation-university-without-language-barriers-1.11379