Automatic Speech Translation at the Threshold

Automatic real-time speech-to-speech (STS) translation is coming of age. Google Translate now has 500 million users a month; Skype Translator is off to an impressive start; and several smaller players are active. Machine translation promises to break down language barriers while creating cultural revolutions and large new global markets.

And yet, Gartner – for the third year running in 2015 – still characterizes STS as "hype." This characterization may be uncharitable in view of recent progress; but the fact remains that none of the participants in this space are generating significant revenue today by selling STS services. Without a clear roadmap to critical mass, startups that seek to grow the STS ecosystem are unlikely to be funded, and VCs are likely to watch and wait.

The gating factors for widespread adoption of speech translation, we believe, are verification, correction, and customization facilities. Users will tolerate the inevitable translation errors if they can catch them and, when necessary, fix them. Without feedback on translation accuracy and a degree of control over the translation process, they are much less likely to rely on automatic systems.

The crucial reliability thus gained will put STS over the usability threshold for serious vertical markets, where customization reflecting each market's special needs can accelerate adoption – and where users will pay for the value received. At the same time, verification/correction facilities will for the first time enable even monolingual users to provide effective feedback for continual improvement of translation engines via machine learning. The resulting expansion of the crowdsourcing base will in turn improve automatic translation for languages with fewer bilingual speakers.

The facilities in question can be engineered for relatively seamless operation and minimal interference with the flow of conversation. Users can decide on the degree of oversight appropriate for the situation.

Verification, correction, and customization facilities will also benefit other natural language processing tasks, such as the handling of queries and commands within Language User Interfaces (LUIs) like Siri, Google Now, and Cortana.

Broadening use cases and creating business models

STS today benefits greatly from major investment by large players like Google and Microsoft/Skype. Until now, however, efforts have understandably been directed at global, and thus general-purpose, consumer markets; and the associated applications are currently free – they are not presently intended as revenue centers. Vertical-market solutions in which STS clearly has monetary value, such as healthcare, the military, law enforcement, language learning, intelligence, and many others, have not yet attracted venture capital. True, there are many major companies with dominant positions in vertical markets where STS could take off, thus giving rise to ecosystems with viable business models. Normally, however, startups must first prove the market; and this has yet to happen.

The main reason for continued hesitation is the perception that translation is still insufficiently reliable for use in vertical markets, and that in fact the translation is least reliable where the need is greatest.

Reliability is indeed crucial. (We purposely stress "*reliability*" rather than "quality" or "accuracy" to connote not only measurable error reduction but the additional and decisive emotional factor of user trust.) Inappropriate translations are obviously unacceptable in critical applications, for example in business or healthcare. Granted, errors are somewhat more tolerable in general-purpose use cases (though we suspect that users' expectations will rise soon); but even here, significant mistranslations can

embarrass, or put a relationship off-course. In both specialized and general-purpose use cases, blind and irreparable errors can damage trust and discourage repeat usage. The risk of distortion or social gaffes may then outweigh even the many obvious benefits of automated translation.

And reliability does indeed decrease when most needed. While reliability of, say, English ↔ German or English ↔ French speech translation is by now often acceptable for simple sentences, it tapers off rapidly with more complex input – just when the need to be understood is the greatest. Likewise, reliability suffers when very different or less populous languages are involved, precisely those for which translation aids are most needed. Problems are well known, for instance, between English and Japanese, languages that differ greatly in structure. For pairs like English ↔ Pashto, or even Japanese ↔ Pashto, the scarcity of bilingual speakers increases the demand for automatic translation and at the same time decreases the effectiveness of present techniques for crowd-sourced improvement of translation (as explained below).

Verification/correction and customization facilities will be key elements in increasing translation reliability and in extending reliable translation to diverse or less populous languages. (Of course, many other elements will contribute to improvement of translation; but the mentioned facilities will play early and decisive roles in increasing user trust, with many beneficial side effects.)

The trust that can speed user adoption largely depends upon such tools. Without them, users can't know when significant errors occur; and even if they did know, they would have no recourse. If speech translation is to progress beyond general-purpose use cases to more serious scenarios and to universal acceptance, trust is essential; but if tools for verification and correction are missing, trust will prove difficult to build. For this reason, we believe, users will come to demand them. We believe that users will pay for translation solutions that provide sufficient verification, correction, and customization, while systems that do not will be treated as commodities, good enough for casual use but not worth paying for.

Beyond their value for building trust, the same verification and correction tools will yield the huge additional benefit of greatly extending the crowdsourcing base for improvement of translation systems by enabling even monolingual users to supply effective feedback for machine learning. This extension will in turn accelerate the improvement of the translation technology itself and its extension to diverse or minor languages.

As these changes take hold, reliability in speech translation will pass the usability threshold for serious vertical markets while at the same time broadening the technology's appeal in general-purpose global markets. The path toward significant revenue will then be cleared, and with it the path to significant investment and rapid maturation of this world-changing technology.

User control of verification tools

While real-time interactive verification and correction is necessary for user confidence, it's also clear that smooth communication is desirable, and that interaction could potentially interfere. To some extent, then, there is a trade-off between verification and correction tools and seamless or frictionless user experience.

However, this potential conflict can be resolved by equipping verification and correction facilities with user controls and built-in intelligence. For example, we can distinguish between pre-verification (checking translations prior to transmission and pronunciation) and post-verification (checking afterward). Post-verification can be made the default mode; and in this mode, conversations can proceed entirely without interruptions, but users can check when and if ready whether the translations were accurate, and can retrieve and repair imperfect translations when necessary. Pre-verification can be triggered only when the confidence score of a translation falls below a threshold preset by the user; or simple interface elements can let users manually activate pre-checking when reliability is paramount, and turn it off when the conversation is routine. In other words, users can decide for themselves what degree of verification is

required for the situation. Today, the vendor decides preemptively; but when users are in control of a translation system, they'll use it more confidently, with more tolerance for occasional interruptions and with reduced tendency to blame the system for errors. System adoption and market growth will be accelerated accordingly.

Verification tools and extension of machine learning

Current efforts to continually improve machine translation via crowd-sourced interactive translation and machine learning are promising, and Google's leadership in this area is impressive.

However, current crowdsourcing solutions face a sharp limitation. To provide useful feedback on translation quality, crowd members must be bilingual: to catch and correct translation errors, they must know both languages. This requirement retards engine improvement, since it excludes most of the world and favors major languages.

Verification and correction tools can be used to elicit effective feedback from even monolingual speakers, thereby dramatically scaling up the "crowd" for crowd-sourced improvements via machine learning. Each correction made by every monolingual user via his or her native language will provide input for continued improvement.

Correction and customization tools in natural language processing

And one more thing: verification/correction and customization facilities will be vital not only in translation but in "Language User Interfaces" (LUIs) such as Siri, Google Now, and Cortana. The option to pre-check commands and queries prior to execution will enhance user confidence and satisfaction, at the same time facilitating feedback to continuously accelerate improvement via machine learning.

Summary

Automatic speech translation is progressing impressively, but commercial development has been held back by the perception that translation remains insufficiently reliable for use in paying vertical markets. Users and investors don't yet trust systems to translate reliably in business, healthcare, and other serious use cases, and even general-purpose consumer use is unnecessarily constrained.

Verification, correction, and customization facilities will be key elements in lifting translation over the reliability threshold for vertical markets as the tools are refined by venture-backed startups and integrated into enterprise-class systems shipped by established software companies. For global general-purpose markets targeted by Google, Microsoft/Skype, and others, seamless integration of the facilities will build user trust and accelerate adoption, and will help extend reliable translation to diverse or less populous languages.

These developments will spur maturation of a world-changing industry just coming into its own and poised at a critical turning point. Its promise is thrilling indeed.

By Faruq Ahmad, in collaboration with Dr. Mark Seligman. Mr. Ahmad is Founding Partner of Palo Alto Capital Advisors, and has been a serial entrepreneur and venture capitalist. He can be reached at faruq@paloaltocap.com. Dr. Seligman is CEO of Spoken Translation Inc., a leading company in the speech-to-speech translation space.

Spoken Translation, Inc.: Technology and Capabilities

Spoken Translation Inc (STI) was incorporated in 2002, with a focus upon automatic speech translation. The company specializes in verification, correction, and customization of real-time translation and of other natural language tasks, such as handling of queries and commands. CEO Dr. Mark Seligman, a recognized expert, has obtained three granted US patents in this area, with a fourth pending.

System reliability

Adoption of speech translation hinges upon trust in the system, whether in serious use cases like business, healthcare, and emergency response or in relatively forgiving scenarios. Trust can be quickly eroded, however, if significant errors are not caught in time. Unfortunately, when monolingual speakers converse via automatic speech translation, neither can know whether the translation is correct without interactive real-time feedback; and even if a mistake were found, there has been no way to correct it.

This blindness and powerlessness remain the Achilles heels of all real-time translation systems today. Until these shortcomings are addressed, speech translation will be slow to build trust. STI's verification, correction, and customization technologies are designed to address them in order to accelerate the field's commercial success.

Feedback on translation quality from even monolingual speakers

Crowd-sourced feedback enabling machine learning is a promising path toward improvement of real-time and other online translation. Google and other large players are making significant investments in this direction. Until now, however, effective feedback has been possible only from users with knowledge of both input and output languages.

Since STI technology enables even monolingual speakers to verify and correct real-time translations, it can greatly expand the crowdsourcing base: every user can provide effective feedback, and translation improvements via machine learning will be accelerated accordingly.

Converser for Healthcare and user control of verification

STI has developed Converser for Healthcare, a product customized for the healthcare market and successfully field tested at Kaiser Permanente. Converser provides a striking demo of real-time interactive speech translation communications between patients and caregivers in English and Spanish. System tools include Reliable Retranslation™, Meaning Cues™, and Translation Shortcuts™.

The system provides an important measure of user control. Users can easily switch between post-verification mode (in which conversations proceed without interruption but each utterance can be verified after transmission) and pre-verification mode (in which verification and correction is enabled prior to transmission for enhanced reliability).

Translation and other natural language processing systems can be configured for seamless operation, in which pre-verification mode is switched on automatically if a confidence score falls below a specified threshold.

API for developers

STI technology can be quickly baked into commercial natural language processing. Please contact us about plans.

Spoken Translation Inc.

Spoken Translation, Inc. is located in Berkeley, CA. CEO Dr. Mark Seligman received his Ph.D. in computational linguistics from UC Berkeley and has been active in R&D for speech translation since the early '90s. At ATR International in Japan he participated in the first international demonstration of this technology. In cooperation with CompuServe, he organized the first successful demonstrations of open-ended spoken language translation. He has since implemented the interactive verification and correction facilities which form the core of STI's intellectual property, along with customization tools and interface elements enabling flexible control of the interaction. Three US patents have been granted to date, and a fourth is pending.

Contacts: mark.seligman@spokentranslation.com or faruq@paloaltocap.com