iTranslate

Will Automation Render Human Language Interpretation Obsolete?

VOIANCE®

a CyraCom International company

Voiance is often asked about the possibility of automated interpretation and translation through advancing technology. We support and embrace the development of machine interpretation solutions as part of our mission to make language services faster and more available to everyone who needs them. This whitepaper explores the history and current status of machine interpretation technology and explores how it may be used today and in the future.

INTRODUCTION

It's been a staple of most science fiction universes for 70 years – the universal translator. First postulated in Murray Leinster's 1945 novella First Contact, the idea of a device to facilitate flawless conversations between languages and cultures has spanned properties like Doctor Who, Star Trek, and The Hitchhiker's Guide to the Galaxy. The methodologies vary but the practical effect is always similar; English-speaking protagonists can discourse in their own language and, through the magic of the universal translator, be heard in the native tongue of their audience. The device works both ways and is seamless, silent, and an incredible asset to intergalactic adventurers everywhere.

Back on planet Earth, the idea of using a computer to bridge the communications gap between the languages of the world may actually predate the electronic computer itself. Separate patents were filed in France and Russia for "mechanical multilingual dictionaries" as far back as 1936. The first public demonstration of Machine Translation (MT) occurred on January 7, 1954 at Georgetown University, in which a computer translated 250 words from Russian into English. The demonstration sparked public interest in the concept and spurred research dollars into the field for the next decade, a trend which lasted until the 1966 release of the Automatic Language Processing Advisory Committee report.

The ALPAC report concluded that perfect automated MT might be impossible and that it was inferior to human translation in "quality, speed, and cost." While the report has since been judged by some to be unfair, it had a chilling effect on research into MT, resulting in funding being pulled and research largely abandoned for close to two decades.

The ALPAC report concluded that perfect automated MT might be impossible, and that it was inferior to human translation in "quality, speed, and cost."

In the modern era, automated interpretation is back in the news, thanks to Skype's recent and impressive Translator demonstration, showing schoolchildren in the US and Mexico successfully conversing back and forth in English and Spanish. Google Translate has premiered its own mobile app, boasting near-instantaneous speech translation in more than 80 languages.

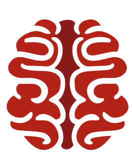
Still, even as technological breakthroughs continue to replace humans with software in a widening variety of industries… there still appears no viable substitute for the real-world professional human interpreter.

These advances are certainly impressive – real-world tests show the software can at least deliver most of the right words, in an order which certainly gives the general idea of the message, if not with any particular grammatical grace. Still, even as technological breakthroughs continue to replace humans with software in a widening variety of industries – cars are slated to be driving themselves in ten years – there still appears no viable substitute for the real-world professional human interpreter. And a true understanding of human speech interpretation, as well as the functionality a computer would need in order to perfectly replicate it, calls into question whether such technology is even possible outside of science fiction.

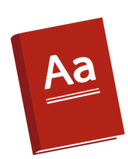
INTERPRETATION AND THE HUMAN MIND

The ability of the human mind to translate (text) and interpret (speech) is often overlooked. It's easy to lose track of the complexity of the task, given that people do it every day and tend to make it look easy, but consider the steps. A person must:

1. Hear a passage in the original language



2. Understand the meaning of both the individual words and phrases, and the passage as a whole



3. Take that thought and determine how to express it in the target language

4. Speak the passage in the target language

Let's break these steps down into further detail to better understand what each entails.

1. Hearing the Passage: At first glance, this step seems too simple to consider. Still, it is noteworthy, as automated attempts to recognize and separate individual spoken words continue to struggle. The difficulties inherent in Siri's attempts to understand and answer spoken questions and Google Voice's transcribing of their customers' Voice Mails into text messages are clear indicators that this, the first step to speech-to-speech interpretation, is still a work in progress for machines.

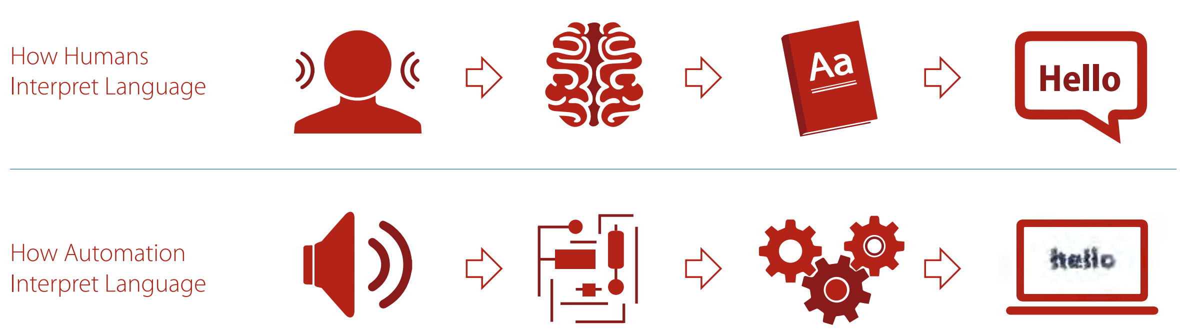
2. Understand the Meaning: Humanity really shines here. The human brain takes in a stream of written or spoken words and, without conscious effort on the part of the interpreter, understands them. To do this, the brain must ascribe a meaning to each individual word, but also understand the complex working relationship between the words and the role each word plays in the overall meaning of the sentence. At times, humans also simultaneously choose from several possible meanings of the word, accounting for context of the sentence, recipient, culture, timing, and more factors.

3. Conversion into Target Language: The next step is to take the passage – now fully understood in its original language – and convert it into the target language, finding words and phrases in the target language which best relay the meaning and intent of the original language. These may or may not be the direct, word-for-word translations of the original words – sometimes a literally correct interpretation does not convey the intended meaning. Me apuñaló por la espalda (Spanish for "he stabbed me in the back"), for example, indicates literal attempted murder, rather than betrayal, as an English speaker would understand it.

4. Speaking the Passage: Lastly, the human interpreter converts the now-translated text into speech in the target language, intuitively drawing upon her knowledge of the language and understanding of the rules which govern it. This involves correct word usage, grammar, pronunciation, and knowing which of several pronunciations of a given word (in the case of heteronyms) is appropriate in a given context.

A human interpreter must perform each of these four steps with diligence and accuracy. If he mishears the speaker, he may produce a wonderful translation of what he thinks was said. The same is true if he hears the speaker's words but misunderstands the message, or if he chooses just one wrong word when converting that meaning into the target language. All the parts must work together seamlessly to produce an accurate, usable, professional interpretation – one misstep and the whole thing falls apart.

THE CHALLENGES OF AUTOMATION



Understanding the deceptively complex task of human interpretation brings the challenge of automating the process into better focus. The Defense Advanced Research Projects Agency (DARPA) has pursued this goal for some time, seeking a near-perfect speech-to-speech interpretation program for field use during military activity. This is not a small problem for DARPA – their parent agency (Defense Department) recently spent nearly $700 million on human interpreters in a single year, so there remains enormous incentive to develop an alternative that doesn't require a paycheck. Despite the enormous benefits and the resources at DARPA's disposal to make this dream a reality, a number of obstacles persist:

1 & 2. Machine Speech Recognition & "Understanding": Just as with a human interpreter, the first step to automated interpretation is to speak a message and have your interpreter correctly identify each spoken word as the one you intended. Humans do this without conscious thought. Teaching a computer to correctly identify the unrestricted speech of an unlimited number of users and convert it to text, though, turns out to be a surprisingly difficult problem - one for which current available solutions remains imperfect at best.

Best practices for accurate automated speech recognition include:

Limiting vocabulary: the smaller the vocabulary used to speak to the program, the easier time the computer has recognizing each word accurately. For example, an automated system can easily distinguish between the 10 digits "zero" through "nine" when spoken but loses accuracy as more words are added to its vocabulary.

Speaker dependence: speech recognition is more accurate when the program interacts with a single human speaker over and over, learning their particular speech pattern. Ideally, the user spends a substantial amount of time speaking program-provided phrases to the computer, allowing the program to "learn" the speaker's voice.

Isolated speech: a computer will recognize spoken words more accurately if the speaker pauses distinctly between each word. The speaker should also avoid "um"s and "ah"s, to avoid confusing the system.

Ideal environment: a quiet, isolated environment without ambient noise will improve speech recognition programs' accuracy.

Unfortunately, none of the above describes what the market for universal translation is truly seeking – a program which allows users to speak using their entire vocabulary, in fluent sentences, wherever they happen to be, using the technology even for the first time, which performs with acceptable accuracy. And the bar for "acceptable" in this field is quite high – how many words out of every 100 can the program miss, without human intervention, before its results are unusable? Speech recognition software developers claim 90-95% accuracy. Is one incorrect word in every 10-20 words acceptable? Depending on which word was missed and what the computer substitutes, the results may be anything from mildly annoying to disastrous or mortifying.

3. Text Translation

Assuming that speech recognition software can correctly identify each word in a spoken message and convert it to text, the heart of any such program remains translation - converting the message from the original language to the target language and capturing the correct meaning. At a glance, this step seems as straightforward as opening a computerized version of, for example, an English-to-Spanish dictionary, looking up each English word, and substituting its Spanish equivalent on the page. This is translation at its most basic, and it has largely been achieved through technology and Internet databases.

However, simple word-for-word translation merely scratches the surface of true, automated interpretation technology. The next edition of the Oxford English Dictionary will reportedly include 645 distinct meanings of the word "run" – which one should an automated translator choose?

The "there/their/they're" distinction, a bane of many human writers of English, seems ready-made to stymie a computer program as well. Likewise, word order and grammar rules do not carry over from one language to another, so even a perfect word-for-word translation program produces less-than-ideal interpreted results.

The next edition of the Oxford English Dictionary will reportedly include 645 distinct meanings of the word "run" – which one should an automated translator choose?

In the past, DARPA and the Army have bypassed these issues using "phraselators" - programs with predetermined English sentences translated by humans into the needed language - but their scope is limited: inputs must be strictly controlled and limited in order to get the right output. Manually writing specifically encoded rules for translating the grammar and structure of one language to another is also considered impractical; it would take hundreds, if not thousands of rules to achieve accuracy from one language to another, and a new set of rules would have to be written for each language pairing – with over 6,000 languages on Earth, this could conceivably require millions of total rules. And any viable solution must also address the homonym/multiple meaning problem; in essence, you must teach a computer to recognize context.

The next step, then, has been Statistical Machine Translation (SMT). The idea is for a computer program to comb vast databases of previously (human) translated documents and to learn patterns and write algorithms based on the choices made by human translators. The program must recognize and code patterns in each individual language to determine things like sentence structure and word order; it can then apply these findings when it comes to translating works into that language. This type of program attempts to work on a sentence-by-sentence basis, using the rules it has learned to place words in order than convey the meaning of the original language, while conforming to the sentence structure and grammatical norms of the target language.

Unfortunately for the sentence-to-sentence approach, different languages require a varying number of sentences to make the same point. A program designed to read human translations and derive the meaning of each sentence may miss the fact that, say, the human interpreter needed two Spanish sentences to match one original English sentence. The computer program may then inaccurately believe that only the first Spanish sentence is the equivalent to the English one, with the second Spanish sentence then being paired with the next English sentence to come, throwing off every subsequent pairing for the rest of the document. SMT also relies on there already being a large volume of human-made translated material for its software to comb and learn from; such material is available for the more commonly spoken languages but may not be for those less popular.

There's also the problem of knowing whether a given translation is accurate without having them all proofread by a human – exactly the kind of tedious, human-capital-intensive activity automation seeks to avoid.

SMT has made good progress over word-for-word translation; still, DARPA plans to go further by teaching their program to understand syntax – the grammatical roles that nouns and verbs play in different sentences. This has led to improvements to word order issues but has yet to produce translation or interpretation indistinguishable from that of a human. There's also the problem of knowing whether a given translation is accurate without having them all proofread by a human – exactly the kind of tedious, human-capital-intensive activity automation seeks to avoid.

DICTIONARY ▸ PHRASELATOR ▸ SMT ▸ SENTENCE TO SENTENCE ▸ SYNTAX

One answer is an IBM-crafted metric called B.L.E.U., which purports to "score" the accuracy of machine translations by comparing them side-by-side with a human translation of the same passage and checking for matching words. This metric is limited though, since it can only measure whether the machine used the same words a human translator chose, not whether the same meaning was conveyed – the true test of an accurate translation. An alternative is to have a human translator review the machine's work, scoring it on how close the machine got to the text's intended meaning, but again, that's a very human-labor-intensive solution; viable, perhaps, for honing the program's accuracy but not for ongoing usage.

4. Text-to-Speech

Assuming all goes well with speech recognition and that the translation program to which it is paired produces an accurate equivalent in the target language, a seamless automated interpretation program must also speak the translation in an understandable way. This requires a number of resources – first among them, an audio file for each word or sound in the target language, from which to compile a thought. Also, where homonyms (sound-alike) words can trip up the translation stage of the process, heteronyms (same spelling, varying pronunciations) become an issue at this point. Should the computer say "PRO-ject" or "pro-JECT"? Upon encountering the text "read," should it play the audio clip for "red" or "reed"? This last step, then, adds another layer of complexity beyond text-to-text translation.

LEGAL LIABILITY

Perhaps the most underappreciated potential concern with a fully automated interpretation program is that of legal liability. The issue manifests itself in several ways worth addressing.

First and foremost, there is potential liability associated with inaccuracy. If your organization relied on a service like Google Translate or Skype Translator, resulting in inappropriate action based from an inaccurate interpretation, where does the liability fall? Too often, the organization using the MT would be held responsible for the error and resulting damage, by virtue of their having relied on/contracted with an unreliable partner. And Google, at least, does not leave this to chance, having spelled out specifically in their terms of service that users must hold them harmless for inaccuracies and that they will accept no liability:

> ***"Business uses of our Services:***
>
> *If you are using our Services on behalf of a business, that business accepts these terms. It will hold harmless and indemnify Google and its affiliates, officers, agents, and employees from any claim, suit or action arising from or related to the use of the Services or violation of these terms, including any liability or expense arising from claims, losses, damages, suits, judgments, litigation costs and attorneys' fees.***"**

This issue could theoretically arise using human interpreters as well; however, a reputable language services provider should have training and performance monitoring which ensures they are providing accurate interpretation results – something not guaranteed from a free automated interpretation program.

Less obvious, from a legal liability standpoint, is the issue of confidentiality. Any agreement between your organization and a human language services vendor would, of course, include confidentiality provisions: does your vendor store data? Where? How is it secured? Do they gain any rights to your data by virtue of their having interpreted it for you? What procedures has the vendor put in place to ensure their agents protect your data?

Do they gain any rights to your data by virtue of their having interpreted it for you?

If a vendor proved incapable of answering these questions or if their answers were less than reassuring, you would likely shop elsewhere. To date, however, online interpretation and translations services do not appear to provide these same safeguards. Google, probably the current biggest name in automated online translation, includes the following in their Terms of Service for using Translate:

> ***"11. Content license from you:***
>
> *11.1 You retain copyright and any other rights you already hold in Content which you submit, post or display on or through, the Services. By submitting, posting or displaying the content you give Google a perpetual, irrevocable, worldwide, royalty-free, and non-exclusive license to reproduce, adapt, modify, translate, publish, publicly perform, publicly display and distribute any Content which you submit, post or display on or through, the Services. This license is for the sole purpose of enabling Google to display, distribute and promote the Services and may be revoked for certain Services as defined in the Additional Terms of those Services."*

Read plainly, anything input into Google Translate remains your property, but you have essentially granted Google license to use it in any manner they wish.

Read plainly, anything input into Google Translate remains your property, but you have essentially granted Google license to use it in any manner they wish. This presents a significant legal problem for any field in which there is an expectation or requirement of confidentiality, e.g., medicine, law, business, and government, as it leaves an organization legally liable for failing to comply with existing confidentiality regulations, having released a client's information beyond their span of control.

WHERE DOES MACHINE INTERPRETATION MAKE SENSE?

The aim of a true universal translator is clear: seamless communication with anyone, any time, worldwide. The financial benefits to medicine, business, and governments (and, by extension, taxpayers) around the world would be substantial, given that the US alone spends hundreds of millions annually on language services. Tourism, commerce, cooperation between the nations of the world – all could see leaps forward with instant, accurate, and affordable automated interpretation. Machine Translation may soon make that dream a reality; it is difficult to tell with certainty *when*, as science has predicted that this technology is right around the corner for decades.

Still, there are many situations where using MT makes a lot of sense today. Skype's recent demonstration showed schoolchildren in different countries communicating in their own languages using the Translator program, illustrating its potential as an effective educational aid. Google's Word Lens feature allows a user to point their smart phone's camera at any text and receive an on-screen translation, aiding any world traveler looking to unravel the mysteries of foreign-language street signs, headlines, and menus; as well as recent immigrants struggling to navigate their new nation of residence. Businesses are also making good use of speech recognition and translation software; automated phone systems already use them to route incoming calls by topic and doing so by language seems like a logical next step. The Army's "phraselator" device illustrates Machine Translation in its ideal environment: repetitive, transactional contacts in which inputs are limited to a set list of phrases and corresponding output can be preprogrammed.

Current MT still has challenges to overcome: speech recognition glitches, imperfect automated word-for-word text translation, and confidentiality measures must improve before the technology can tackle complex, professional interactions. In the meantime, human interpreters handle these same tasks daily. So, at least until Machine Translation realizes its full potential, human interpreters will continue to bridge the gap between the nations, languages, and cultures of the world.

Voiance is a division of CyraCom International, Inc. ("The Company"), the leading provider of language interpreting services. Whether in-person or via phone, mobile app, or written text, Voiance bridges communication gaps for organizations that need rapid access to language assistance. In business since 1995, the Company services thousands of clients throughout the US, including many Fortune 500 companies.

Voiance's translation services operate under two certifications: International Standards Organization (ISO) Quality Management Systems (9001:2008) and Translation Services (17100:2015). These complementary guidelines help ensure quality, accuracy, and continuous improvement.

Contact Voiance today to discuss how we can improve your language services program.

Phone: (844) 727-6739 | info@voiance.com | www.voiance.com
Mailing Address: Voiance | 5780 North Swan Road | Tucson, Arizona 85718