

Across the Language Barrier

Translation devices are getting better at making speech and text understandable in different languages.

“THE GREATEST OBSTACLE to international understanding is the barrier of language,” wrote British scholar and author Christopher Dawson in November 1957, believing that relying on live, human translators to accurately capture and reflect a speaker’s meaning, inflection, and emotion was too great of a challenge to overcome. More than 60 years later, Dawson’s theory may finally be proven outdated, thanks to the development of powerful, portable real-time translation devices.

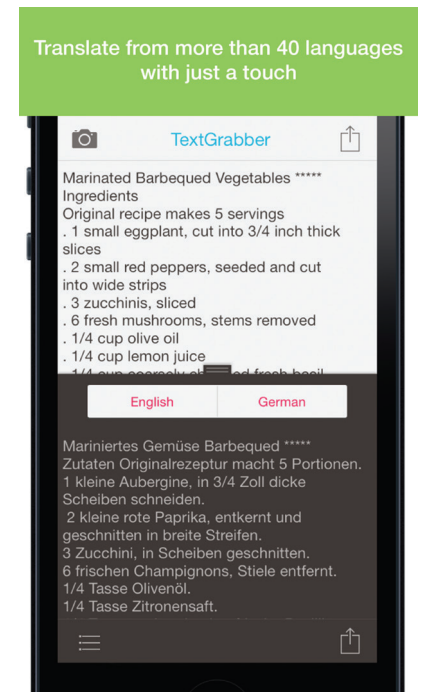
The convergence of natural language processing technology, machine learning algorithms, and powerful portable chipsets has led to the development of new devices and applications that allow real-time, two-way translation of speech and text. Language translation devices are capable of listening to an audio source in one language, translating what is being said into another language, and then translating a response back into the original language.

About the size of a small smartphone, most standalone translation devices are equipped with a microphone (or an array of microphones) to capture speakers’ voices, a speaker or set of speakers to allow the device to “speak” a translation, and a screen to display text translations. Typically, audio data is captured by the microphones, processed using a natural language processing engine mated to an online language database located either in the cloud or on the device itself, and then the translation is output to the speakers or the screen. Standalone devices, with their dedicated translation engines and small portable form factors, are generally viewed as being more powerful and convenient than accessing a smartphone translation application. Further, many of these devices offer the ability



to access translation databases stored locally on the device or access them in the cloud, allowing their use in areas with limited wireless connectivity.

Instead of trying to translate speech using complex rules based on syntax, grammar, and semantics, these language processing algorithms employ machine learning and statistical modeling. These initial models are trained on huge databases of parallel texts, or documents that are translated into several different languages, such as speeches to the United Nations, famous works of literature, or even multinational marketing and sales materials. The algorithms identify matching phrases across sources and measure how often and where words occur in a given phrase in both languages, which allows translators to account for differences in syntax and structure across languages. This data is then used to construct statistical models that link



phrases in one language to phrases in the second, which allows for accurate and fast translation.

In practice, this means devices can translate between languages more quickly than ever before by using such modeling. Incorporating high-powered processors, quality microphones, and speakers into the device, a person can carry on a real-time, two-way conversation with someone who speaks an entirely different language. These devices represent a significant increase in accuracy and functionality above manual, text-based translation applications such as Google Translate.

The advances in technology have not gone unnoticed, as the market for language translation devices is projected to reach \$191 million annually by 2024, up from slightly more than \$90 million annually in 2018, according to data from Research & Markets. Much of the activity is due to

the growth in international travel and tourism, particularly from residents of countries where English language proficiency is relatively low.

For example, countries such as Japan, China, and Brazil feature a strong middle class with the means to travel internationally. Yet, these countries each are ranked “low” on the 2018 Education First English Proficiency Index (EPI), reflecting the challenges many travelers have when leaving their home country.

The ideal solution is for citizens to learn to speak multiple languages, according to Howie Berman, executive director of The American Council on the Teaching of Foreign Languages. “Our position has always been that technology is a complementary piece to the language learning process,” Berman says. “I think language really depends a lot, it’s not just on what you say, but how you say it. And, I think translation devices really do fail to pick up on a lot of the cultural cues.”

However, the casual traveler may not have the time or inclination to become proficient in a new language in preparation for a tourist trip or event, like the 2020 Olympic Games in Japan, or the 2020 FIFA World Cup scheduled to be held in Qatar. For these one-off trips, Berman says, “We certainly don’t expect someone going to the Olympics to enroll in multiple classes right before they go; we realize that’s not feasible for everyone.” Regarding modern translation devices, Berman says, “We think they’re valuable tools, but we see them for what they are, as complementary tools to the classroom experience.”

Still, the use of machine learning will help translators become better at understanding nuance, regional dialects, and tone. As algorithms are trained on voice data containing these characteristics of everyday speech, the accuracy and intelligence of the models will improve over time, particularly with translations between languages that do not feature similar structures or character sets.

One device that addresses these concerns is Pocketalk, a standalone translation device developed and marketed by Japanese software company Sourcenext Corp., which the company says can translate between 74 languages. Pocketalk has shipped globally more than 600,000 units of the \$230

The use of machine learning will help translators become better at understanding nuance, regional dialects, and tone.

device since its debut in 2017, capturing nearly 96% of the global translation device market, according to April 2019 data from analyst firm BCN Retail.

“Pocketalk was created to connect cultures and create experiences for people that do not speak the same language, and can and should be used for both business and leisure,” says Joe Miller, general manager and product lead for Pocketalk. Miller says Pocketalk’s translation engines can recognize local dialects, dialect nuances, slang, and accents. “The voice translation will use an accent when speaking back the translation, not a robotic voice,” Miller says.

However, like other devices designed to support live, multiple-way conversations, Pocketalk relies on a connection to the Internet to access its online language database and translation engine. Devices that feature a limited number of languages often can store these databases on the device, but devices that support dozens of languages generally require a persistent connection to a cloud database. While Pocketalk works on 4G cellular connections, devices such as Birgus’ Two Way Language Translator or the ODDO AI pocket translator require the use of a Wi-Fi connection, and will not work using only a cellular connection.

Devices that require a Wi-Fi connection may not be suitable for travelers who spend a lot of time interacting with people outside of formal indoor settings, as they may not be able to access a reliable Wi-Fi signal. That drawback is less of an issue for translating devices designed for the international business user community, who utilize translation devices to conduct real-time business meetings and seminars that require two or more languages to be translated.

“Through our research we found that there was a need for a translator that is optimal for professional uses and can support multiple people easily conversing at the same time,” says Andrew Ochoa, founder and CEO of Waverly Labs, creator of the Ambassador, a small over-the-ear translation device that can support up to 20 languages and 42 dialects, but which requires the use of a companion iOS or Android mobile application paired to a smartphone to function. “Whether someone is participating in one-on-one conversation, a multi-person meeting, or larger conference setting, Ambassador allows them to easily listen and communicate with their colleagues and teams.”

The Ambassador incorporates a series of microphones, and combines the input with speech recognition neural networks, in order to capture speech clearly. The system also utilizes cloud-based machine translation engines built on translation models that incorporate local accents and dialects, allowing Waverly Labs to use machine learning to tune the accuracy of their devices based on regional parameters.

When traveling, not all communication is verbal. Fujitsu also offers a portable standalone translation device similar to Pocketalk, called Arrows Hello, which also includes a camera that can capture images, such as signs and menus that include foreign characters, and then display the translations of those text-based materials on its screen. Similarly, optical character recognition (OCR) technology company ABBYY offers a consumer-focused mobile app called TextGrabber that can “read” text or QR codes in more than 60 languages, then translate the words or phrases to a different target language while retaining the appropriate syntax and meaning, according to Bruce Orcutt, the company’s vice president of product marketing.

“ABBYY’s an OCR company, so you can imagine our bias towards converting everything text that’s possible,” Orcutt says. The TextGrabber app, he says, “uses multiple technologies that have evolved and developed to ultimately identify text, and then we use our OCR technology once we have identified the text.” TextGrabber em-

ACM Member News

DESIGNING COMMUNITIES WHERE MINDS MEET



Amy Bruckman, a professor and senior associate chair in the School of Interactive Computing at

the Georgia Institute of Technology, says her interest in computer science began when she took computer science classes in high school.

Bruckman went on to earn her undergraduate degree in physics from Harvard University, her master of science degree from the Massachusetts Institute of Technology (MIT) Media Lab Interactive Cinema Group, and her Ph.D. from the MIT Media Lab's Epistemology and Learning Group.

After receiving her doctorate, Bruckman joined the faculty at the Georgia Institute of Technology, where she has remained ever since.

Her research interests include social computing, collaboration, social movements, content moderation, conspiratorial ideation, and Internet research ethics.

"I am also writing a book," Bruckman continues. "It is called *Should You Believe Wikipedia? Understanding Knowledge and Community on the Internet*," and "is based on a lot of the things I teach in my Design of Online Communities class, which I have been teaching since 1997. I am hoping to share some of what I have learned in teaching the class in the book."

She anticipates the book will be published next year, by Cambridge University Press.

Bruckman says she has "one fun dream," which is to facilitate communication between people who hold radically different political views, so they can come to understand one another better. "I don't know how to do that, but maybe I will have invented some kind of Internet game or discussion forum where that can happen."

—John Delaney



plays machine learning algorithms to identify text within an image, applies OCR to capture that text, then applies a logic engine to clean up syntax and character misreads, such as being able to discern whether a character is a zero or the letter "O," based on context.

While TextGrabber currently does not include any functionality for capturing voice or video to aid in real-time translation, its OCR translation technology is incorporated into solutions from Microtek, Panasonic, Ricoh, Sharp, and others. Orcutt believes that in the future, devices that can handle any type of media, including audio, moving video, images, and text, will become commonplace.

"If you look at the younger generations, [those] digital first generations, they have no problem navigating these tools, as they're part of their ecosystem," Orcutt says. "And I think with the 2020 Olympics coming up in Japan, there'll be a tremendous amount of innovation in this area to help. I know the Japanese government is interested in making the Japanese market more easily navigated by tourists to make the Olympic experience better."

Clearly, technology developments in machine learning have led to devices that can provide accurate, real-time translations for people attending large, multinational-focused events such as the Olympics. Berman, however, hopes

these technical achievements may spur people to take the next step and actually try to learn another language to fully understand its nuances, via a combination of technology and traditional classroom instruction.

"I think it's wonderful that these devices and these tools are elevating the status of language," Berman says. "We think [translation devices] are valuable tools, but we see them as complementary tools to the classroom [learning] experience." **C**

Further Reading

Brown, Peter F. et al.

"A statistical approach to language translation." COLING (1988). <https://www.semanticscholar.org/paper/A-statistical-approach-to-language-translation-Brown-Cocke/2166fa493a8c6e40f7f8562d15712dd3c75f03df>

Wenniger, Gideon Maillette de Buy.

"Aligning the foundations of hierarchical statistical machine translation." (2016). <https://www.semanticscholar.org/paper/Aligning-the-foundations-of-hierarchical-machine-Wenniger/de12e7ecf32523ac9b480d3dab052ec5b43ebef9>

What Buyers need to know about speech translation devices: <https://www.youtube.com/watch?v=LUVNcp2xQQM>

Keith Kirkpatrick is principal of 4K Research & Consulting, LLC, based in Lynbrook, NY, USA.

© 2020 ACM 0001-0782/20/3 \$15.00