Meta takes us a step closer to Star Trek’s universal translator

As a result, the AI translators we have today support an impressive number of languages in text, but things are complicated when it comes to translating speech. There are cascading systems that simply do this trick in stages. An utterance is first converted to text just as it would be in any dictation service. Then comes text-to-text translation, and finally the resulting text in the target language is synthesized into speech. Because errors accumulate at each of those stages, the performance you get this way is usually poor, and it doesn’t work in real time.

A few systems that can translate speech-to-speech directly do exist, but in most cases they only translate into English and not in the opposite way. Your foreign language interlocutor can say something to you in one of the languages supported by tools like Google’s AudioPaLM, and they will translate that to English speech, but you can’t have a conversation going both ways.

So, to pull off the Star Trek universal translator thing Meta’s interviewees dreamt about, the Seamless team started with sorting out the data scarcity problem. And they did it in a quite creative way.

Building a universal language

Warren Weaver, a mathematician and pioneer of machine translation, argued in 1949 that there might be a yet

undiscovered universal language working as a common base of human communication. This common base of all our communication was exactly what the Seamless team went for in its search for data more than 70 years later. Weaver’s universal language turned out to be math—more precisely, multidimensional vectors.

Machines do not understand words as humans do. To make sense of them, they need to first turn them into sequences of numbers that represent their meaning. Those sequences of numbers are numerical vectors that are termed word embeddings. When you vectorize tens of millions of documents this way, you’ll end up with a huge multidimensional space where words with similar meaning that often go together, like “tea” and “coffee,” are placed close to each other. When you vectorize aligned text in two languages like those European Parliament proceedings, you end up with two separate vector spaces, and then you can run a neural net to learn how those two spaces map onto each other.

https://arstechnica.com/science/2025/01/meta-takes-us-a-step-closer-to-star-treks-universal-translator/