Multilingual December 2016, Community Lives
Is a Universal Language Possible?
By Jeannette Stewart

According to Noam Chomsky, linguist and political thinker, “A language is not just words. It’s a culture, a tradition, a unification of a community, a whole history that creates what a community is. It’s all embodied in a language.” So, what does language mean for a global community? A universal, global language unifying and embodying all of us? Given the diversity of human life is that even possible? There are some possible precedents. First there was maybe Proto-Indo-European and/or Babel, then Hellenistic Koiné, then Latin, then Esperanto. What also of Arabic, Hebrew or Mandarin? Not to mention C, HTML, or Python? After all, computers talk to each other in 1s and 0s regardless of the language used to program them and they span the globe. It seems that wherever languages are used, the desire for some form of universal language is identified as a means of circumventing the one-to-one translation process. The idea of a bridge (a koiné language) connecting a number of languages, understandable to a large population, does indeed have a strong appeal, especially when a goal of globalism is real-time, multilingual communication. Does a universal language make sense in today’s network-connected world?

Language has many functions. Indeed, in the present era, we may say that it has many applications. Linguists work to identify the ways and means we use languages to transmit information; interact and commune with each other; simply converse in a massive variety of modes. This lies at the heart of the question in hand. We do not have a universal means of communicating with each other quite simply because we do not have a universal topic to discuss: we have millions. This is excellent news for translators and localizers. Perhaps not such good news for those hoping that computer-assisted translation will be a magic bullet for cross-cultural communications. Yet the idea persists and with a growing appreciation of what characterizes a global community, it is still an idea under investigation.

As Latin fell into decline in the post-Renaissance world of European letters, many thinkers sought to replace its abilities to express all manner of subjects to the literati in a widely understood form. Mathematicians Rene Descartes and Gottfried Leibniz attempted to formulate a means of constructing a language capable of expressing conceptual thoughts. In England, John Wilkins, amongst others, sought to facilitate trade and communication between international scholars using a system of ‘real characters’, symbols that constituted a lingua franca. According to Wikipedia, “Whereas a vernacular language is used as a native language in a community, a lingua franca is used beyond the boundaries of its original community and is used as a second language for communication between groups.” This underlines an important distinction between different uses that languages are put to. Set in the context of the present day, we find ourselves with a heightened awareness of minority, endangered languages, yet we also seek to communicate well beyond the confines of our own communities.

In 2001, Professor Abram de Swaan of the University of Amsterdam brought an intriguing and powerful approach to understanding how power and languages are connected in the global
community in his book ‘Words of the World: The Global Language System.’ His accomplishments as a social scientist enabled him to detail how a multilingual world can also be described in hierarchical terms that expose the uneven field upon which languages compete for dominance. In his model, English occupies a ‘hypercentralized’ position, whereas other languages exist more diffusely from central to peripheral positions. In the translation community, we work in this arena on a daily basis. There have, of course, been critics of de Swaan’s ideas in the academic community, but the work has been highly influential in furthering our understanding of the way in which communication can facilitate human affairs globally.

Theoretical approaches to specifying how a universal language works are essential to understanding how the global, multilingual community might operate using a single or dominant language. But how might this work in practice? As mentioned above, thinkers in the seventeenth century were interested in using signs and symbols to communicate. This is still an idea that is being explored with the unlikely sounding Lovers Communication System (LoCoS) devised by Yukio Ota, Professor of Design at Tama Art University in Japan. Ota is world-famous for his design of the green running man used to mark exit doors in millions of buildings across the world. With our proliferating use of emojis and their incorporation in Unicode, this is hardly surprising. But their use thus far is largely confined to text messages and websites. They do, however, represent text to varying degrees and it is premature to say just what their future is. That said, if a picture is truly worth a thousand words, then they surely must have a bright future. When actor Kyle MacLachlan was asked to explain the plot of the film Dune on Twitter, he managed to describe the entire movie with 41 emoji characters (see photo). (NOTE to editor: http://www.theverge.com/2016/8/18/12516674/kyle-maclachlan-dune-movie-emoji-quiz link for the Dune tweet.)

The emoji “language” is already recognizable universally and this is due to the Unicode Consortium, which has embraced the new language, and is diligently defining and approving new emoji characters. Every new version of Unicode includes recommendations for implementation but companies are free to represent the emoji whichever way they wish, thus growing the range of expressions. With growth comes diversification, confusion and misunderstandings. With representations now covering multiple skin tones and occupations having a female variant, Unicode is doing a spectacular job in providing creative solutions. For example, the gender-diverse emoji for occupations is a combination of the standard “man” and “woman” emoji with a second emoji to represent the occupation. These are joined together by a special invisible character called the “zero-width joiner” (ZWJ). Platforms that support the new emoji, recognize the ZWJ and display a single emoji while others will display two separate ones. The ability to create new emoji brings its own problems, mainly fragmentation and inability to include them in the official Unicode version. For example, Twitter has a “pirate flag”, Windows has added “ninja cats” and WhatsApp has an Olympic Rings emoji, which in other platforms is shown as five plain circles. The potential for confusion and misrepresentation across different platforms can only be avoided by sticking to the official Unicode version. As the emoji language grows and increases its expressions, its universal nature is what appeals to people.

I wonder, though, just what the implications of transcribing the contents of a
communication made up of emojis are for the language community? Will we need specialists to translate the plot of Dune in Twitter messages?

With cultural and commercial imperatives driving the world’s need for instant and universal communication, it’s hardly surprising that many place great faith in technology to provide universal, workable solutions to common problems. This has, of course, transformed the language community with localization and globalization becoming essential components in delivering products and services in an international marketplace. The academic community has responded and is now offering a range of localization courses to an eager body of students, who in turn are much sought-after by corporate businesses.

The internet has always strongly gravitated towards community structures. We all use phrases like the ‘online community,’ the ‘virtual community,’ the ‘hacking community,’ the ‘user community,’ the ‘developer community’ and so on. Product development methodologies like Agile allow us to manage huge multilingual tasks in a coherent and productive manner even if the community working together is a globally-distributed one. At least, that’s the theory.

Software tools like Translation Memories (TMs) have enabled large-scale projects to achieve their goals faster and cheaper while still relying on the human expertise of language services professionals that we all want to see elevated to its rightful status. After all, the ATA skillfully and properly put the White House in its place when a call was made in 2009 for bigger and better automated translation. In a deftly-worded response to President Obama, ATA President Dr. Jiri Steskel asserted that “translation software and qualified human translators are vital to your goal of achieving language security.

Today all the leading proponents of computer translation recognize that human beings will always be essential, no matter how sophisticated translation programs become.” I doubt any language professional would disagree and this tends to suggest that there is assuredly no place for a universal language in our community. However, the pace of change at the cutting edge of tech is still running at a blistering rate. Welcome to the brave new worlds of the Internet of Things and Machine Learning.

Picking up on Chomsky’s idea that all aspects of a community are embodied in its language, can we say the same for a community’s technology? The ancient Greek word ἀπόστασις (apósasti, meaning ‘afar’), which we find in ‘telephone,’ ‘television,’ ‘telecommunication’ and so on, bridges enormous distances. These devices shrink our world, but they enlarge our communities. The Internet of Things promises to connect us to an even greater degree if we are to believe the hype. We are promised that just as our phones pack enormous computing power into a hand-held device, billions of gadgets will be similarly empowered. A recent report identified ‘implementation problems’ as a barrier to progress in achieving this ultra-ambitious internet-connected network of devices. Implementation in what respect? Business and tech analysts will cite security and privacy as massive headaches or achieving robust and reliable connectivity in a massively heterogeneous networked world. But what if your device speaks one language and you speak another? Will we have to localize smart fridges? The answer has to be yes. If Siri, Apple’s Virtual Assistant (VA) is available in a growing variety of languages; if PayPal’s services are now available in over 200 languages; if Amazon has operations in at least
international locations; then to paraphrase H. G. Wells, "I've seen the future and it's multilingual."

So what exactly is powering this hugely diverse yet intimately connected world of tech? The computing community, like the language community, is made of many other diverse communities, of which one is the Artificial Intelligence (AI) community. In turn, it too is made up of many very-varied communities. AI, which used to be regarded by more mainstream computing communities as exotic to say the least. That, however, has changed dramatically and AI is now truly mainstream itself. AI has many areas of application, but one that is of particular interest to the language community is Natural Language Processing. In particular, Machine Learning (ML) is being applied to endow computers with the capability of 'understanding' texts and, taking a further leap, of 'translating' them.

With a field that draws input from computer science, cognitive psychology, neurolinguistics, data science and numerous theories of education, it is no wonder that numerous different approaches are taken to automating language acquisition. It would be counter-productive to even attempt to generalize efforts in the field. However, two approaches to training a computer to learn how to translate a language are worth a very brief examination. These are: rule-based systems and statistical systems. We should note that neither of these are cognitive approaches. They involve processing.

Rule-based systems rely upon a set of syntactic and orthographic conventions that are used to analyze the content of a source text, which provides the input to be generated into the target language. But the problems with using this approach are obvious. Word order, for example, is anything but universally the same. Indeed, the notion of core grammar just does relate to the real world of diverse language families, not to mention accommodating isolates like Basque. Clearly the problems can be overcome such that there is a way of connecting languages in pairs, but that for the present relies upon the hard work of the poor old human linguist.

The other approach involves the use of statistical processing based on bilingual text corpora. Google Translate is the perfect example of this approach which harnesses raw processing power to detect patterns of equivalence in language pairs. Almost all of the texts that are mined in this way are the product of human translators in the first place and this is what gives proponents of this approach confidence that the quality of the output target will be of satisfactory quality. Another benefit of this approach is that it is responsive to use with new language pairs and this gives some researchers hope that a monolingual text corpus, the engine of a universal translator, is a future possibility.

I am so fortunate to have worked with some incredibly talented translators and interpreters. I can now add to that by working with some superb engineers and computer professionals. As I look at the community of communities that now make up the language industry, it's breathtakingly diverse, yet has achieved so much in so many places around the world. I am also lucky to live in times when our technologies are developing so quickly that we can barely keep up. If computers are not already everywhere, they soon will be. And what of our silicon friends who speak at light-speed in 1s and 0s? Will they ever achieve consciousness as some researchers believe? Here's a parting thought from AI researcher, Alan Stewart, who...
is working on neural networks, on the topic of computer cognition. “I am optimistic about the future capabilities of computers and by that I mean that raw power and sophisticated logic will create amazing technology, but unless there are some startling breakthroughs, it will still fall short of nature’s biological capabilities.” However, he speculates that with the learning capabilities that computers are being given, it’s not outwith possibility that they will begin to look for more efficient ways to achieve the tasks we ask them to do. That’s one of the products of learning algorithms. At the recent DEFCON in Las Vegas, a Cyber Grand Challenge was staged that pitted two computer systems against each other with the aim of discovering weaknesses in the opposing systems. The results fell short of present human standards, but this is just the beginning. With computers able to learn large amounts of material at high-speed, a new communications paradigm is a strong possibility. For example, is it possible that computers will actually create their own language? I know it sounds ridiculously far-fetched, but there was a time not so long ago that we scoffed at even quick-and-dirty MT. That universal language may still be out there in the future, but will we be able to understand it?