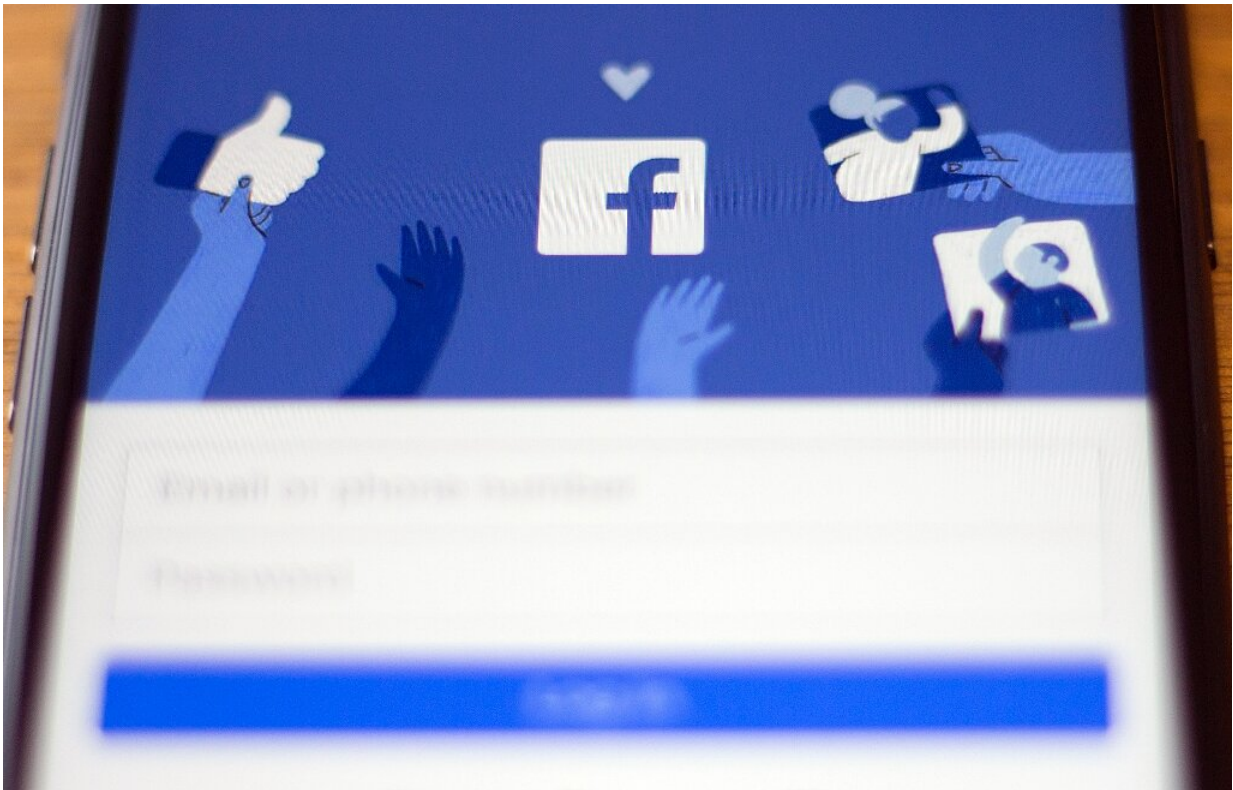# Facebook researchers use maths for better translations

October 13 2019, by Laurent Barthelemy

Social networks such as Facebook are seeking efficient automatic translation for all of the world's languages and artificial intelligence may hold the answer

Designers of machine translation tools still mostly rely on dictionaries to make a foreign language understandable. But now there is a new way: numbers.

Facebook researchers say rendering words into figures and exploiting mathematical similarities between languages is a promising avenue—even if a universal communicator a la Star Trek remains a distant dream.

Powerful automatic translation is a big priority for internet giants. Allowing as many people as possible worldwide to communicate is not just an altruistic goal, but also good business.

Facebook, Google and Microsoft as well as Russia's Yandex, China's Baidu and others are constantly seeking to improve their translation tools.

Facebook has artificial intelligence experts on the job at one of its research labs in Paris.

Up to 200 languages are currently used on Facebook, said Antoine Bordes, European co-director of fundamental AI research for the social network.

Automatic translation is currently based on having large databases of identical texts in both languages to work from. But for many language pairs there just aren't enough such parallel texts.

That's why researchers have been looking for another method, like the system developed by Facebook which creates a mathematical representation for words.

Each word becomes a "vector" in a space of several hundred dimensions. Words that have close associations in the spoken language also find themselves close to each other in this vector space.

From Basque to Amazonian?

"For example, if you take the words 'cat' and 'dog', semantically, they are words that describe a similar thing, so they will be extremely close together physically" in the vector space, said Guillaume Lample, one of the system's designers.

"If you take words like Madrid, London, Paris, which are European capital cities, it's the same idea."

These language maps can then be linked to one another using algorithms—at first roughly, but eventually becoming more refined, until entire phrases can be matched without too many errors.

Lample said results are already promising.

For the language pair of English-Romanian, Facebook's current machine translation system is "equal or maybe a bit worse" than the word vector system, said Lample.

But for the rarer language pair of English-Urdu, where Facebook's traditional system doesn't have many bilingual texts to reference, the word vector system is already superior, he said.

But could the method allow translation from, say, Basque into the language of an Amazonian tribe?

In theory, yes, said Lample, but in practice a large body of written texts are needed to map the language, something lacking in Amazonian tribal languages.

"If you have just tens of thousands of phrases, it won't work. You need several hundreds of thousands," he said.

'Holy Grail'

Experts at France's CNRS national scientific centre said the approach Lample has taken for Facebook could produce useful results, even if it doesn't result in perfect translations.

Thierry Poibeau of CNRS's Lattice laboratory, which also does research into machine translation, called the word vector approach "a conceptual revolution".

He said "translating without parallel data"—dictionaries or versions of the same documents in both languages—"is something of the Holy Grail" of machine translation.

"But the question is what level of performance can be expected" from the word vector method, said Poibeau.

The method "can give an idea of the original text" but the capability for a good translation every time remains unproven.

Francois Yvon, a researcher at CNRS's Computer Science Laboratory for Mechanics and Engineering Sciences, said "the linking of languages is much more difficult" when they are far removed from one another.

"The manner of denoting concepts in Chinese is completely different from French," he added.

However even imperfect translations can be useful, said Yvon, and could prove sufficient to track hate speech, a major priority for Facebook.

© 2019 AFP