What do they mean when they say neural machine translation and why should I care?

Mikel L. Forcada,
Universitat d'Alacant,

The last years have witnessed an increase in interest in a new machine translation (MT) paradigm: *neural machine translation* (NMT),[1] pioneered years ago by researchers in Spain.[2] NMT, part of the wider field of *artificial intelligence*, is displacing its predecessor, statistical MT (SMT). Many on-line systems (Google, Microsoft, etc.) have turned *neural* and there are even *neural-born* systems (DeepL). Companies such as Amazon, Google, Facebook, etc., are building their own NMT technologies, and translation service providers (TSPs) are gradually adopting NMT. In this communication I will explain —in accessible terms— how NMT systems work, how they relate (remotely) to the functioning of the nervous system, how they learn to translate from existing translations, and how NMT differs from other MT technologies. I will help the audience to decipher the jargon of NMT (*distributed representations*, *embeddings*, *vectors*, *layers*, *weights*, *encoders*, *decoders*, etc.) and will build upon these concepts to explain the architecture of NMT systems. As with any other technology, it is crucial that translators and translation students, are able to grasp the basics of this new technology and become aware of what they can expect from it, through a critical reading of the hype created by some of its proponents. This communication is intended to contribute to those goals. It will also analyze how the results (and *errors*) of NMT are different from those produced by SMT or rule-based MT, and will discuss the impact of the computational and corpus requirements of NMT systems on translation work-flow.

[1] Forcada, M. L. (2017). Making sense of neural machine translation. Translation Spaces, 6(2), 291-309.

[2] Castaño, A., and Casacuberta, F. (1997). A connectionist approach to machine translation. In *Fifth European Conference on Speech Communication and Technology EUROSPEECH'97*; Forcada, M. L.,and Ñeco, R. P. (1997, Juny). Recursive hetero-associative memories for translation. In International Work-Conference on Artificial Neural Networks (pp. 453-462). Springer, Berlin, Heidelberg;