intelligence; therefore, it is not strange to talk about →*Neural Machine Translation* as an artificial intelligence application.

Artificial Neural Network

A network of →*Artificial Neurons*, connected in a specific way to accomplish a specific computational task. Connections have strengths or *weights* that can be *learned* and regulate how the state or *activation* of an artificial neuron is affected by the *activations* of neighboring neurons (or the value of external inputs), so that the whole network has the desired behaviour.

Artificial Neuron

An artificial neuron (vaguely inspired on a natural neuron) is a very simple computational units. Its state or *activation* depends on the stimuli received from other artificial neurons (or external inputs) through its connections.

Attention-Based Neural Machine Translation

A kind of →*Neural Machine Translation* using a *learned* mechanism called *attention* to obtain information from selected parts of the source sentence every time a word of the target sentence is produced.

Black-Box approach

According to Wikipedia, "a black box is a device, system or object which can be viewed in terms of its inputs and outputs [...], without any knowledge of its internal workings. For instance, using an MT system in a black-box approach means that one only has access to the output it produces for the source text it is fed. See →*Glass-Box Approach*.

Controlled Language

A subset of a natural language obtained by restricting its grammar and vocabulary in order to make it easier to process by computers; for example, by removing ambiguities and other natural-language features that make machine translation difficult.

Convolutional encoder

A specific kind of →*Encoder* using a →*Convolutional neural network*.

Convolutional neural network

When the same artificial neural network is repeatedly applied to every position (for instance, every word) in the input data and its neighbouring positions (for example, neighbouring words) to obtain a new level of context-informed representation for each position. The resulting representations may be processed again in a similar way to generate another level of *deeper* representations (hence *deep learning*) of the original data.

Corpus-based machine translation Machine translation that uses knowledge *learned* from very large parallel corpora containing source-language segments and their target-language translations. Examples: →*Statistical Machine Translation*, →*Neural Machine Translation*. As opposed to →*Rule-Based Machine Translation*.

Decoder In →*Neural Machine Translation*, an →*Artificial Neural Network* that computes the likelihood that each word in the vocabulary is the next word in the target segment, by looking at the preceding words in the target segments, at one or more *representations* of the previous steps it has taken, and, optionally, using *attention* (see →*Attention-based Neural Machine Translation*), at the representations produced by the →*Encoder* at each position of the source segment.

Encoder In →*Neural Machine Translation*, an →*Artificial Neural Network* that reads a source segment and *encodes* it into a series of →*Representations* which will later be processed by a →*Decoder* to generate a target segment.

Entropy In information theory, a measure of the amount of information that is missing before a message is received, therefore, a measure of *uncertainty*. When the entropy is measured in *bits*, it is the minimum number of yes/no questions needed to fully determine (and therefore encode) the contents of the message. Entropy therefore relates to the likelihood of every possible message: when all messages are equally likely, the entropy is maximum. When a single message is the only one that is likely to appear, entropy approaches zero.

Glass-Box approach A *glass box* is a device, system or object for which one does not only have access to the output it produces for the inputs it is fed, but one whose internals (such as intermediate results) can be viewed (but usually not altered). For instance, using an MT system in a glass-box approach means that one, in addition to the output it produces for the source text it is fed, has access to internal or intermediate stages of how it has been computed. See →*Black-Box Approach*.

Interactive Translation Prediction Also called *Interactive Machine Translation*, a machine translation system that, instead of producing in one go the translation of a whole source segment, provides professional translators with machine-translated suggestions on how to continue the target segment they are typing.

Language model A language model (usually a *probabilistic* language model) assigns a likelihood or score to a sequence of words, usually a sentence. They are commonly used in →*Statistical Machine Translation* to favour fluent translations.

***N*-gram** A contiguous sequence of *N* items (for example, words) as seen in a sample of sequences (for example, a corpus of sentences). Statistics about word *N*-grams gathered from a text corpus may be used, for instance, to build probabilistic →*Language models*.

Neural Machine Translation A kind of →*Corpus-Based Machine Translation* using →*Artificial Neural Networks* trained on large parallel corpora. Most Neural Machine Translation systems consist of an →Encoder and a →Decoder, each of which is a specialized artificial neural networks.

Open-Source A piece of software is *open-source* (also called *free* software or *free/open-source* software) if it is distributed under a license that allows users to run the software for any purpose as well as to study, modify, and distribute it and any modified versions. Typical free/open-source licenses are the Apache license, the MIT license, the BSD 3-clause license, the GPL license, and two Creative Commons licenses (CC-BY and CC-BY-SA).

Phrase-Based Machine Translation [also *Phrase-Based Statistical Machine Translation, PBSMT*] →*Statistical machine translation* using *phrase* pairs, that is, correspondences between sequences of one or more contiguous words, such as (*statistical machine translation, traduction automatique statistique*) as building blocks.

Recurrent encoder A specific kind of →*Encoder* using a →*Recurrent neural network*.

Recurrent neural network An →*Artificial neural network* that at each cycle or time step *t* receives, among its inputs, part of the outputs it produced in the previous cycle or time step *t–1*. Recurrent neural networks are capable of processing a sequence of inputs of any length and produce an output at each step that depends on what has been processed until now.

Representation In →*Artificial Neural Networks*, the vector made up of the activations of a selected group (usually called a *layer*) of →*Artificial neurons*, written as a →*Vector*. Representations of

	similar items of knowledge (words, sentences, concepts) are usually mathematically related.
Rule-Based Machine Translation	Machine translation that works by applying grammar rules and dictionaries written by experts, as opposed to →*Corpus-Based Machine Translation*.
Statistical Machine Translation	A type of →*Corpus-Based Machine Translation* that generates translations based on probabilistic models obtained by performing statistics on parallel corpora. Statistical machine translation systems assign an approximate probability or likelihood to many possible translations of the source segment and then pick up the most likely. See also →*Phrase-Based Machine Translation*.f
Tokenization	Segmentation of source text in *tokens*, usually (but not always) words or similar units, before sending to a machine translation system: *We won't go to Ben's party* → We + wo + n't + go + to + Ben + 's + party.
Vector	An ordered, fixed-size list of real numbers such as "(–0.73. 0.31, 0.76, –0.02)".
Word Embedding	A representation of a word as a →*Vector*. Used in →*Neural Machine Translation*. Word embeddings are usually learned from large monolingual texts. For similar words, embeddings are also mathematically similar.