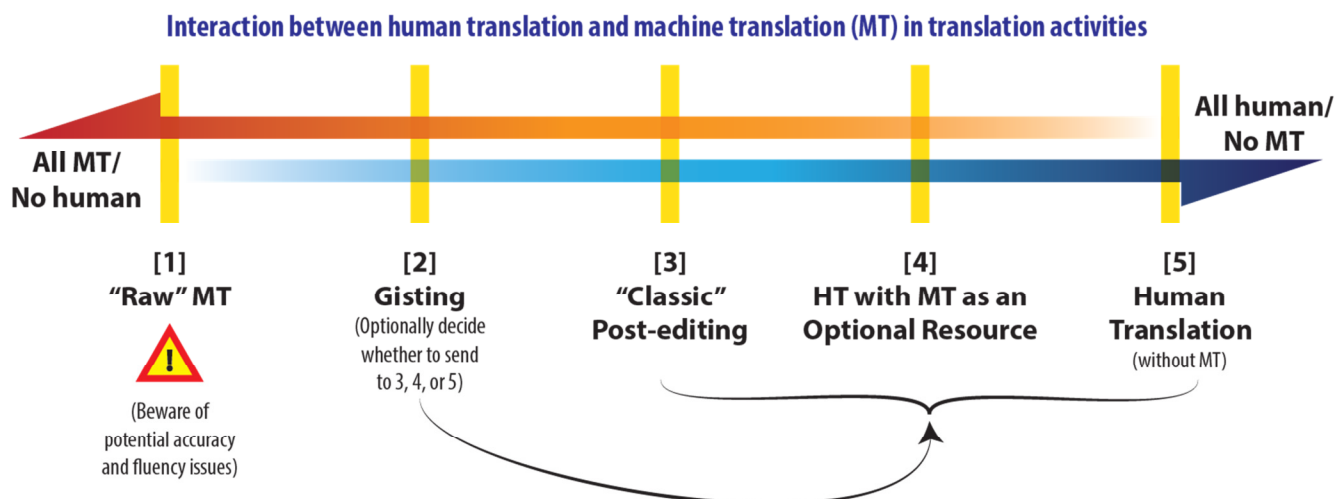


A spectrum from all MT at the left end to all HT at the right end.

Core acronyms: MT = machine translation; HT = human translation



Explanation of the Five Points along the Spectrum

Raw Machine Translation with No Human Post-processing (point 1 on the spectrum)

In this diagram, "raw MT" refers to output from a machine translation system (rule-based, statistical, or hybrid) that has not been modified in any way by a human. All types of automatic post-processing of raw machine translation are included in this point on the spectrum. Of course, much direct and indirect human involvement, including the use of corpora of human translations, is typically part of developing and/or training of MT systems, but humans are not involved once a source text is put in. Raw MT is very fast but often inaccurate and lacking in fluency. Its usefulness depends on purpose.

Human Translation (points 4 and 5 on the spectrum)

In this diagram, "human translation" is an activity that *begins with a source text* and a set of specifications (implicit or explicit) and results in a target text. If the specifications are appropriate and the translation meets them, it is a high-quality translation. During the translation process, the translator typically has access to various resources, such as terminology and translation-memory lookup, that can be used or ignored at the discretion of the human. The main difference between points 4 and 5 on the spectrum is whether machine translation of some or all segments is available to the translator as an optional resource. Use of technology is required in point 4 and allowed in point 5.

Post-editing (point 3 on the spectrum)

In this diagram, post-editing is a human activity that *begins with a raw machine translation* and instructions to the human, including how much effort to put into improving the raw MT. Typically, either light (i.e. partial) or full post-editing (i.e. making it equivalent to HT) is specified. For some specifications, post-editing is not cost effective, and human translation should be done instead.

Gisting (point 2 on the spectrum)

In this diagram, gisting is a human activity *that begins with raw MT*, like post-editing, and results in the human forming a general idea (a “gist”) of the content of the document for some purpose. The person doing the gisting is typically not a translator and does not even need to understand the source language. Gisting ranges from only determining the relevance of a text to understanding well enough to summarize (limited by the accuracy of the raw MT), without looking at the source text. A different human activity than gisting as used in this diagram, sometimes called “summary translation”, is conducted by a human translator. It begins with the source text, and results in a written summary in the target language. So gisting is a *monolingual* human activity while summary translation is *bilingual*.

Sometimes gisting is done only for personal benefit. Other times, it is used as the basis for a human judgement as to whether the document should be sent on for further action, such as post-editing or human translation. This decision process is called triage and is analogous to triage in emergency medical environments where someone decides which patients are first sent to doctors for treatment. Triage is often done quickly in the mind without separating gisting from decision making. Automatic summarization and categorization are not gisting and thus are attached to point 1 of the spectrum.

Triage is not translation. It was used in the United States during the Cold War to decide which Russian research articles should be sent to human translators and studied further by scientists. Triage is currently used in the European Commission by many non-translators who decide, on the basis of a raw machine translation of that document, whether to request a professional human translation of a particular document. A new type of triage involving end-users is being used in the commercial sector. Raw machine translations of short documents, such as technical support entries in a database, are made available online almost immediately after they are authored. End-users who have trouble understanding a particular document may have the option of making immediate corrections (post-editing) or requesting an improved version, which may result in the document being sent out for more processing. In this third type of triage, documents that are useful as raw machine translation or are never consulted do not use up valuable human resources for further post-editing or translation.

How to Use the Spectrum

Once the specifications for a particular translation activity are made explicit, this diagram can be used by a language services adviser to explain to the requester which point on the spectrum is appropriate for that particular translation activity, in order to best accomplish the objectives of the requester's organization. The spectrum should not be seen as a workflow. It supports collaborative clarification of translation requirements by the requester and the provider and should be used with the standard translation parameters found in ASTM F2575 and other standards, along with templates for typical categories of translation activities, as the basis for developing structured translation specifications. The specifications are then matched up by a *language services adviser* with a point on the spectrum.

If disagreement arises as to which point on the spectrum is appropriate, that is, whether raw MT is sufficient in a particular situation or whether post-editing or human translation is more effective, a translation quality metric can be used to determine which option best meets the specifications.