

EIGHT TYPES OF TRANSLATION TECHNOLOGY

Computers are used in many aspects of modern translation (particularly of technical texts). The following information explains the eight main types of computer-aided translation tools and their use in translation environments. This handout describes these functions, as grouped in the chart below. On the reverse side of this page is a sample English-Spanish bitext to which the examples make reference. (Note: a segment is a coherent piece of text larger than a term, usually a sentence.)

INFRASTRUCTURE		
	TERM LEVEL	SEGMENT LEVEL
BEFORE TRANSLATION	<ul style="list-style-type: none"> • Term candidate extraction • Terminology research 	<ul style="list-style-type: none"> • New text segmentation, previous source-target text alignment, and indexing
DURING TRANSLATION	<ul style="list-style-type: none"> • Automatic terminology lookup 	<ul style="list-style-type: none"> • Translation memory lookup • Machine translation
AFTER TRANSLATION	<ul style="list-style-type: none"> • Terminology consistency check and non-allowed terminology check 	<ul style="list-style-type: none"> • Missing segment detection and format and grammar checks

TRANSLATION WORKFLOW AND BILLING MANAGEMENT

Organization of the eight translation tool functions.

1. **Infrastructure.** The infrastructure for a translation environment is not necessarily translation-specific, but the importance of infrastructure becomes even more important in multilingual situations. Elements of the infrastructure need to be as integrated as possible, both among themselves and with the actual translation process. The elements of the infrastructure are:

- Document creation/management system
- Terminology database
- Telecommunications (intranet/Internet, e-mail, FTP, web browsing, etc.)

2. **Term-level before translation: Term candidate extraction and terminology research.** Term candidate extraction and terminology research are used to determine what words might be candidates for inclusion in a term base. After a source-language term is identified, by candidate extraction or some other process, terminology research is needed to find an appropriate term in the target language to designate the concept. Terminology research can draw on many resources, including the Internet and multilingual text databases. As an example, if we assume that the sentences in the bitext on the next page were part of a large text, and that *thermal layer* were not already in the termbase an extraction tool should propose it as a candidate term, even if both *thermal* and *layer* were already in the termbase as individual words. Thus term candidate extraction goes beyond what a spell checker can do by identifying candidates for new multi-word terms.

3. **Term-level during translation: Automatic terminology lookup.** Automatic terminology lookup, though vastly simpler, could be thought of as the term level equivalent of machine translation. For example, in the bitext on the next page the words *thermocline* and *thermal layer* might be considered terms that should always be translated consistently. Automatic terminology lookup would display the preferred target language term (*gradiente térmico* and *capa térmica* in these cases) without the translator having to look the terms up manually. As each segment of source receives the focus, preferred target language terms are displayed and the human translator can quickly incorporate them into the target text without risk of misspelling. Automatic terminology lookup supports terminological consistency for all text types.

4. **Term-level after translation: Terminology consistency check and non-allowed terminology check.** Terminology consistency checkers verify consistent use of terminology after a translation has been completed; i.e., they make sure that each term is translated consistently, wherever it occurs. For example, if the preferred term for *thermocline* is *gradiente térmico* and a human translator, for whatever reason, returns *termoclino*, a terminology consistency checker would detect this inconsistent use and flag the term for human attention. Non-allowed terminology checkers flag terms which are not allowed (as in the case of deprecated terms) and bring them to the attention of a human.

	SOURCE TEXT	TARGET TEXT
1.	He heard the captains discussing the absence of a thermocline.	Oyó que los capitanes comentaban la ausencia de gradiente térmico.
2.	Mancusco explained that it was not unusual for the area, particularly after violent storms.	Mancusco explicó que no era extraño en la zona, particularmente después de tormentas violentas.
3.	They agreed that it was unfortunate.	Convinieron en que era mala suerte.
4.	A thermal layer would have helped their evasion.	Una capa térmica hubiera facilitado la evasión

The sample bitext (given above) is taken from the English original of Tom Clancy's *The Hunt for Red October* and its Spanish translation. A bitext is a set of texts consisting of a source text (English in this case) and target text (Spanish here) which have been *aligned* so that each segment of source text corresponds to a segment of target text.

5. Segment-level before translation: New text segmentation, previous source-target text alignment, and indexing.

The preparation of an aligned, indexed source-target bitext is vital for the correct functioning of translation memory tools if previously translated text is to be *leveraged* (re-used). Indexed bitexts are also useful for terminology research.

6. Segment-level during translation: Translation memory look-up and machine translation. Automatic translation memory (TM) lookup applies primarily to revisions of previously translated texts and requires an indexed bi-text to function. TM lookup compares new versions of texts with the TM database and automatically recalls those segments which have not changed significantly, allowing them to be leveraged. For example, if the third sentence above were completely rewritten but the surrounding sentences were unchanged, TM lookup could process the text and automatically place retrieved translations of the unchanged sentences in the output file and return the changed sentence to the translator who could supply a translation. For minor revisions of previously translated documents, TM lookup can provide enormous productivity increases.

Machine translation takes a source text and algorithmically processes it to return a translation in the target language. Machine translation parses a sentence of source text, identifying words and relationships, selects target language terms, arranges those words in target language word order and inflects them. MT typically is used for controlled language texts from a narrow domain and requires some post-editing where publication quality output is required. MT systems often allow users to modify their dictionaries. The following is raw (unedited) MT output in Spanish of the English source given above (in this case *thermocline* was returned untranslated since it was not in the system's dictionary):

Él oyó a los capitanes que discuten la ausencia de un *thermocline*. Mancusco explicó que no era raro para el área, particularmente después de las tormentas violentas. Ellos estaban de acuerdo que era infortunado. Una capa termal habría ayudado su evasión.

7. Segment-level after translation: Missing segment detection and format and grammar checks. These functions are closely related to #4. They check for missing segments, correct grammar, and correct retention of formatting. For example, if the following translation of the English passage in the bitext were received from a translator, a missing segment detection tool would let the user know that something was missing (the second sentence):

Oyó que los capitanes comentaban la ausencia de gradiente térmico. Convinieron en que era mala suerte. Una capa térmica hubiera facilitado la evasión.

8. Translation workflow and billing management. While workflow management is not directly part of translation, it is extremely important for tracking the progress of translation projects. Workflow management tools keep track of the location of outsourced translations and their due dates, text modifications, translation priorities, revision dates, and so forth. The larger the text and the more texts in process, the more important these features become since the logistics of dealing with all the variables which may influence a project are compounded with size. Billing management also becomes increasingly important as the size of projects increases. Ideally both parts of this function should be integrated with one another.