# Glossary of translation tool types

<table>
<thead>
<tr>
<th>Tool type</th>
<th>Description</th>
<th>French equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active terminology recognition tools</td>
<td>Active terminology recognition (ATR) tools automatically analyze texts in electronic format in order to identify occurrences of terms or other items in that text that are also present in a terminology database. When they identify occurrences of known terms, they may highlight them in the text, propose equivalents stored in the database for insertion in the text by the user, or automatically replace occurrences of these known terms with their equivalents from the database. ATR tools must be combined with a terminology base (usually stored in a terminology management system) and are often integrated with translation memory systems. Some examples of ATR tools include the pre-translation function in LogiTerm and the TermBase Agent in MultiTrans.</td>
<td>Traducteurs de vocabulaire</td>
</tr>
<tr>
<td>Bilingual concordancers</td>
<td>Bilingual concordancers allow users to search for occurrences of character strings (i.e. sequences of characters) in bitexts (i.e. original texts and their translations that have been aligned and are displayed either side-by-side or one above the other). They usually allow users to search in one or both languages, and offer advanced searching features (e.g. Boolean operators, wildcards). These tools can help users to find, study and compare various occurrences of the character string, and to identify and/or evaluate potential translations of lexical units, phrases, sentences or even structures. Some examples of bilingual concordancers include BeeText Find, ParaConc, the bitext search function of LogiTerm and TransSearch.</td>
<td>Concordanciers bilingues</td>
</tr>
</tbody>
</table>
Bitext aligners

Bitext aligners are used to create bitexts (i.e. to break down original (source) texts and their translations (target texts) into smaller segments and then to match corresponding source and target segments). These tools generally work on a sentence level, and match sentences according to their relative lengths, their place in the text, and sometimes their contents. Since these formal criteria and the nature of texts do not always allow for perfect alignment, most tools also offer functions to help users correct the alignment manually. Bitexts can be used to help in analyzing translations and translation techniques; they can be searched using bilingual concordancers and are also the starting point for the creation of translation memories (cf. the entry for translation memory systems). Examples of bitext aligners include the bitext creation function in LogiTerm, the TextBase Aligner in MultiTrans, the aligner in Fusion Translate, and SDL Trados WinAlign. Some bilingual concordancers, such as ParaConc, also include alignment functions.

Corpora

Corpora are collections of electronic texts that have been assembled to assist users in studying language and its use. They are generally designed to provide a representative sample that gives an overview of a certain type of language (e.g. in a particular register, region, field, or type of text). These text collections may assist users in determining how lexical units are generally used or combined. Corpora are generally searched using either monolingual concordancers (e.g. in the case of monolingual or comparable corpora) or bilingual concordancers (in the case of parallel corpora, also called bitext corpora). Some well-known corpora that can be consulted online include the British National Corpus and Frantext.
Electronic dictionaries are becoming more and more popular alternatives to traditional paper dictionaries. These resources, which may be monolingual, bilingual or multilingual, may be available on CD-ROM or online. They often offer fast, easy and flexible access to the content of dictionary entries to help translators and other users find the information they need. It is important not to confuse these dictionaries with term banks, which, while also usually available online, generally have different purposes and organization. It is also important to note that not all electronic dictionaries are of the same quality; it is particularly important to be cautious when using free online dictionaries. The *Oxford English Dictionary* and the *Dictionnaires de l’Académie française* are very well-known dictionaries that can be consulted online (the latter through the *Centre national de resources textuelles et lexicales*). Others that you may find useful include the *Longman Dictionary of Contemporary English Online*, the *Random House Webster’s Unabridged*, the *Oxford-Hachette* and the *Nouveau Petit Robert* (all installed in the Writing Centre).

Localization tools

Localization is a task that involves the translation and adaptation of a Web page, software application or other product to a particular linguistic and cultural community. Because the task often involves working with complex computer coding, many participants are generally involved in the process. A number of localization tools have been created to assist translators and other localization professionals in managing the complexities of the task (including managing workflow, providing accurate word counts and estimates, separating the textual components to be translated from computer code that must be preserved, and managing terminology and previously translated texts or versions of texts). In addition to dedicated localization tools, translation memory systems and terminology management systems are very useful for many localization projects. Localization tools include Catalyst, Passolo, and WebBudget, which is installed in the Writing Centre.
Machine translation (MT) systems are unlike all of the other tools described in this glossary, because rather than assisting a human translator or other language professional in his or her work, MT systems take charge of the entire process of translating texts. However, this doesn’t mean that language professionals have no role to play. When MT systems are used, humans are most often involved in the revision (called *post-editing*) of the target text produced by the MT system to ensure that it is correct and adequate for its intended use. In some cases, humans may also be able to adjust the way the MT system works (e.g. by adding to or modifying the dictionaries it uses) or by preparing documents in such a way that they can be translated as successfully as possible by the system (called *pre-editing*). MT systems are most useful when source texts can be carefully prepared to be easily translated by the system (e.g. by clarifying ambiguous expressions, using short sentences), and/or when the target text is intended purely to assist in comprehension (and not, for example, for publication). There are different underlying techniques used by MT systems. Some try to imitate the ways in which humans process language (e.g. using grammar rules), while others operate using statistical probabilities or by taking examples of previously translated text as models. Just as human translators may produce slightly different versions of a target text, so will different MT systems. It is important not to mix up the short form *MT* for *machine translation* with *TM*, which stands for *translation memory*. You will find tools such as Systran and Reverso installed in the Writing Centre; a number of other systems, such as Babelfish and Google Translate, are also available online for free. (It is important to be particularly cautious when using free online tools, as they obviously do not offer the same kinds of advantages as locally installed tools that can be adapted for use on specific texts!)
Monolingual concordancers

Monolingual concordancers are computer tools that help in the analysis of corpora. They are used primarily to find and display occurrences of character strings (i.e. sequences of characters) in corpora. They usually offer advanced searching features (e.g. Boolean operators, wildcards). Some present the occurrences retrieved in key-word-in-context (KWIC) format, which displays each occurrence on a separate line, with the character string the user searched for displayed in the centre. This presentation is designed to make comparing multiple occurrences easier and more efficient, particularly as most tools also allow the occurrences to be sorted using various criteria. Analyzing occurrences of character strings can help users to evaluate how words and phrases are used and combined. Many concordancers also offer additional corpus analysis functions (e.g. that can make lists of all of the word forms present in corpora and their frequencies, to help identify particularly pertinent items in a collection of texts). Examples of monolingual concordancers available in the Writing Centre include WordSmith Tools, and the full-text search feature of LogiTerm. Some other concordancers can be used and/or downloaded online for free; these include TextStat, WebConc, WebCorp, AntConc and Corsis.

Other Office tools

The Microsoft Office Suite offers a number of software applications to assist with tasks such as creating, editing and managing presentations (PowerPoint), spreadsheets (Excel) and databases (Access). Translators, writers and revisers may use these tools to access documents for translation and to store and manage terminology, client information and other data. Professors may be interested in using such tools to prepare lectures or conference presentations, to store research data or to calculate grades.
Search engines

The best-known Internet search engine is Google, although there are others (including Yahoo!, Alta Vista, and Ask.com). Search engines analyze the contents of Web pages and create lists of occurrences of word forms found in pages or their URLs, in the information that Web page creators provide about their pages (the pages’ metadata), or in links to pages. This analysis facilitates and accelerates searching online using key words. Search engines also use sophisticated calculations to rank the results of searches in an attempt to present the pages that are most likely to be helpful to a user at the top of the results list. Many search engines also offer a number of specialized searching functions, and each search engine has its own particular syntax that must be learned in order to optimize searches. A different type of search tool, called a meta search engine (e.g. Dogpile), carries out searches in a number of search engines at once, and then synthesizes the results.

Term banks

Term banks are resources that assist in research on terms used in specialized language, and are particularly useful for example for specialized or technical translation and technical writing. Usually bi- or multilingual, term banks are collections of term records, i.e. highly structured entries in databases that store data about concepts that are important in specialized fields (e.g. terms, their equivalents in other languages, definitions, contexts, sources, and observations). This concept-based structure and specialized orientation differentiates term banks from most electronic dictionaries, which are generally organized by lexical item. Two of the best-known Canadian term banks are TERMIUM® and the Grand dictionnaire terminologique (GDT). A well-known European term bank is IATE (Inter-Active Terminology for Europe), formerly known as Eurodicautom.
Term extractors

Term extractors are computer tools that analyze texts in electronic format and identify candidate terms (i.e. lexical units that appear likely to be terms in a specific domain). Term extractors use a variety of methods to find candidate terms, all of them calling upon formal criteria such as frequency analyses or structures of word combinations. Note that this automated software does not produce perfect results. Some actual terms contained in the text may be overlooked, while some candidates that are not actually terms may be proposed. Therefore, the output of a term extractor must be verified by a language professional. Examples of term extractors include SDL Trados MultiTerm Extract and TermoStat. Translation environments such as MultiTrans, LogiTerm and Fusion Translate also include term extraction functions.

Terminology management systems

Terminology management systems (TMSs) are tools similar to generic database management systems, but which are designed specifically to assist translators and other language professionals in storing and managing terminological data (e.g. terms, equivalents, domains, definitions, contexts, and sources). TMSs allow users to create, store, manage and search their own term records for items they feel will be useful in future work. They often suggest or even impose record structures to help users store various kinds of terminological data, and also offer a certain number of search functions to help users find the records they need quickly and easily. TMSs are often a more powerful alternative to more general Office tools (e.g. word processor documents with tables, spreadsheets or databases) for storing terminology. Another advantage is that TMSs can sometimes work in conjunction with an active terminology recognition tool and/or translation memory system as part of a larger translation environment. It is important to note the difference between TMSs (usually used to store personal records or records for a fairly small group of users) and term banks (which are generally organization-wide or even public or commercial products such as TERMIUM® or the GDT). Terminology management systems include SDL Trados MultiTerm and BeeText Term. Translation environments such as MultiTrans, LogiTerm and Fusion Translate also include terminology management functions.
In CERTT, we use the term *translation environment* (sometimes abbreviated TenT for *translation environment tool*) to refer to systems that include a number of different tools for translators in one integrated package. These environments are usually centred around *translation memory systems* or similar tools, but usually also include *bitext aligners*, *terminology management systems*, *term extractors*, *active terminology recognition tools* and *bilingual* and/or *multilingual concordancing* functions, among other tools. Some even allow for integration of *machine translation* for some purposes. Translation environments installed in the Writing Centre include SDL Trados, Fusion Translate, LogiTerm, MultiTrans, WordFast and OmegaT.
Translation memory (TM) systems are designed to save translators time and effort in translating documents that contain repetitions (either internal, within the same text, or external, in other similar documents). TM systems store segments of texts that have been translated, accompanied by their translations, usually in a type of database called a translation memory. (These matched segments, sometimes called translation units, can be created automatically as a user translates, or can be assembled using existing source and target texts matched using a bitext aligner.) Most TM systems link directly to text editors (e.g. word processors), so that they can be used while a translator is working on a translation as he or she normally would. When a translator is working on a text in which a segment is similar or identical to one that has already been translated, the TM system can automatically suggest the previous translation for re-use. The translator can then decide whether this translation is appropriate for use in the new text. If so, he or she can simply insert it in the new translation, or can edit it as necessary and then insert it. If the suggestion is not appropriate, the translator can simply reject it and translate him- or herself. Translation memory (TM) systems should not be confused with machine translation (MT) systems; translation memories allow humans to recycle segments of previous human translations, while MT systems carry out translations automatically and most often rely on humans for editing after the fact. While these two types of tools may be integrated in some translation processes or even in a single translation environment, they function in quite different ways. Translation memory systems form the core of translation environments such as Déjà Vu, Fusion Translate and SDL Trados (specifically the Translator’s Workbench tool). (The latter two tools are installed in the Writing Centre.) Similar tools are also found in the MultiTrans and LogiTerm translation environments.
Web tools
The category of Web tools is a general one, including various types of tools useful for language professionals that can be used online. Some of these are difficult to fit into classical tool classes (e.g. Diatopix, a tool that allows a user to do a number of Web searches limited by region simultaneously, and creates a graph to compare the results, which can be helpful in identifying whether a term or expression is more common in one region than another). Others (e.g. the library catalogue, ORBIS) are more general tools that can nevertheless be useful for translators and others in the language industry. In CERTT, this category is distinguished from electronic dictionaries and term banks, which — while also generally available online — form their own classes of tools.

Windows XP
Windows XP is a computer operating system, a collection of software programs that manages computer resources and coordinates the operation of programs and the management of files and computer-human interaction. Windows functions covered in CERTT that may be useful to translators include creating and extracting files from compressed (i.e. zipped) folders, and using keyboard shortcuts.

Word processors
Word processors are software programs that help users to enter, edit, format and save text documents. Most also offer additional functions that can help translators, writers and revisers to compare or revise documents, to save information in different formats or layouts (e.g. as tables), and to convert files into different file formats. Translation memory systems and other translation tools often interact with word processors to provide assistance to translators directly in the word processor environment.