

COMPARATIVE STUDY OF MODERN CAT TOOLS: SMARTCAT VS MATECAT

The concept of automated translation is shaped by a wide range of definitions, which may lead to a certain confusion. Some use the term “automated” or “automatic” translation to refer to all kinds of translation tools, including machine translation [3]. However, these two notions should be differentiated as they represent distinct concepts that can sometimes conflict with one another. Machine translation (MT) is defined as the process of translating content from one language to another without any human involvement [1]. In other words, it is a fully automated process, which relies solely on algorithms and computational models. Conversely, automated translation typically refers to computer-assisted tools that aid human translators with various features to enhance the efficiency of translation. They allow a human translator to edit, create, store, and manage their translation. CAT tools were designed to help translators to do their job more quickly and efficiently, allowing human translators to “lead the way”.

The idea of CAT tools or translation-memory software emerged in the 60s of the 20th century. The practical implementation of computer-assisted software is typically associated with the concept of “translator’s workstations” or “workbenches”, introduced by Martin Kay in his seminal paper [2]. This concept builds upon the idea of a “Translator’s Amanuensis”, a device consisting of a mainframe computer linked to terminals via wires. This collaborative translation environment provided human translators with individual workplaces allowing them to share mutual translation assets efficiently.

Modern CAT tools have evolved significantly compared to their early predecessors in terms of functionality, user interface, availability and integration with other technologies. They incorporate a range of components, including advanced translation memory, terminology management, quality assurance features, integration with machine translation engines, capabilities for cloud-based collaboration etc.

One of the key components of any CAT tool is translation memory. Translation memory stores previously translated text segments along with their original counterparts, allowing translators to reuse them when encountering similar or identical context again. So, it can be defined as a database that stores segments of text, such as sentences, paragraphs, or headings, along with their corresponding translations for the reuse in future projects. This eliminates the necessity of translating similar sentences from scratch, which saves time and cognitive efforts for human translators, while enhancing consistency, efficiency, and accuracy. The CAT tool breaks the source text into segments and aligns each segment with the segments stored in translation memory, looking for full and partial matches, which in terms of CAT are called *exact* and *fuzzy matches* correspondingly. The translator can either accept the match or replace it with their own translation. Segments which found no matches in translation memory are translated by a translator manually or with the help of integrated MT engine.

Term bases or translation glossaries are another essential component common for most CAT tools. They are databases, where terms, – individual words or phrases, – are stored alongside with their translation and possible explanation. Term bases enhance the accuracy of translation ensuring consistent usage of domain-specific terminology across multiple projects.

Quality assurance features serve as a text-editing software that identifies errors in translation and help to eliminate them. In addition to spellchecking, QA tools can identify missing tags, inconsistent or unapproved terminology, character mismatches etc.

Depending on some particular CAT tool, this software can operate with various file formats, cover a wide range of languages, is usually integrated with an MT engine, offers a collaborative translation environment, allowing to run team projects while sharing common translation resources.

The **goal** of our study is to perform a comprehensive analysis of two popular cloud-based CAT tools, widely used by both professional and novice translators.

MateCat (Machine Translation Enhanced Computer Assisted Translation) is a 3-year research project (2011–2014) funded by the European Union's Seventh Framework Programme for research, technological development and demonstration. It offers a wide array of functions including public and private translation memories, concordance search, revision capabilities and project management. Matecat is a free tool supporting over 200 languages and offering work with 80 file formats.

It has a modern user-friendly interface, starting with the project creation form available directly from the home page www.matecat.com. Matecat is accessible from any location with internet connection using either Google Chrome or Safari. Translators can use MateCat anonymously or log in to have the expanded range of features. Compared to other CAT tools, Matecat has certain benefits. Apart from private translation memories that can either be uploaded from personal computer and attached to the translation project or activated from the list of TMs stored in the profile, this application offers translation matches from Public TM called My Memory Collaborative TM. This public database of human translations developed for Matecat project is constantly updated with translations performed by Matecat users (unless they choose not to share their work). My Memory Collaborative TM ensures that translators have access to a vast, continually evolving resource providing more matches than any other CAT tool.

Another advantage of Matecat is free machine translation engine – Modern MT – developed by Matecat parent's company. ModernMT is a cutting-edge technology positioned as the leading machine translation system for enterprise needs [6]. It checks previous translations for similar sentences and adjusts its output based on past work. Matecat can also work in Revision mode, where one can edit translation providing information about each issue, while the program will be generating a detailed quality report based on the edits. Matecat users have a wide range of management options suitable for team work. Project owners are provided with a splitting option for distributing a large-scale translation job between multiple translators. Alternatively, Matecat allows project managers to set up projects and assign them to teams of translators, which can be created from a management panel.

Additionally, comprehensive guidelines for MateCat can be accessed from the home page. They provide detailed information about all Matecat features, presented clearly and complemented with images and video tutorials.

One of the drawbacks of Matecat performance is that it can be slow during the final stage of project creation. Since the tool is free, projects are typically placed in a queue with the indication of the number of projects which need to be processed before the current one.

Similar to Matecat, Smarcat is a cloud-based application, positioning itself as “all-in-one” translation platform that connects businesses and translation professionals while offering additional services like payment automation. The Smartcat software was initially developed by ABBYY Language Solutions, a linguistic service provider, between 2012 and 2015 as an in-house solution for the CAT tool [5].

It supports 281+ languages and 40+ file formats, from Microsoft Office and hypertext to subtitles and images. There are three main types of accounts in Smartcat: freelancer, translation company, and end customer. The main benefit of Smartcat is its global reach for translation services. By creating a freelancer profile and filling in the necessary information – such as education, work experience, translation samples, certificates, language pairs, services and rates etc. – freelancers can advertise their services and significantly expand the customer base. However, in practice with 150,000 translators registered on the platform, it can be challenging to appear at the top of search results.

Generally speaking, Smartcat has a more complex interface compared to Matecat, which might confuse an inexperienced user and require additional time to fully master its functionality. The process of project creation is relatively simple, involving two main steps: uploading a file for translation and selecting the language pair. Additionally, users can upload linguistic assets such as translation memories, reference files and glossaries. They can also configure advanced settings related to machine translation, translation memory and glossary, quality assurance settings etc.

Smartcat integrates eight industry-leading translation engines to provide instant translations: Google Statistical and Neural translation engine, Microsoft Translator – (also, statistical and neural, depending on the language pair.), Yandex, Baidu Translate API, DeepL, Amazon Translate, and ModernMT. However, only Yandex translation engine is unlimited and free. Additionally, users need to be aware that their translations will be sent to the machine translation provider in order to improve the output. Other machine translation engines impose limits on the number of translations, which can be increased for an extra fee.

Smartcat positions itself as an ideal collaborative environment, which enables managing large projects with a team of translators. However, this option may not be available in a freelancer's account or might require a paid version. Smartcat tool is equipped with a quality assurance feature, which operates within the editing pane marking segments with potential issues using an exclamation mark. It should be noted that no revision capabilities have been identified in a free demo version, making Matecat a more attractive tool in this regard.

A distinctive feature of Smartcat is its capacity to translate video and audio transcripts. By importing a file into a project, Smartcat automatically parses it into subtitles with timecodes. This feature ensures that subtitles are precisely synchronized with the corresponding audio upon upload [4].

Another key consideration for potential users is that the access to certain tools may require additional steps such as contacting the support and changing the type of account, which may complicate the overall workflow. Moreover, the range of options in a free demo version is limited compared to the paid versions.

In conclusion, both Matecat and Smartcat provide powerful cloud-based translation solutions, offering an extensive range of features including translation memories, glossaries, integration with machine translation and quality assurance tools. The choice between these platforms will depend on specific needs and the level of experience. Matecat is particularly well-suited for novice translators offering a clear and user-friendly interface while providing all essential CAT features to gain experience without overwhelming functional complexity. Smartcat, on the other

hand, provides a wider range of options and is suitable for professional translators, experienced in various translation tasks. However, accessing certain features in Smartcat might require an additional fee.

References

1. G2 Technology glossary. <https://www.g2.com/glossary/machine-translation-definition>.
2. Martin Kay. *The Proper Place of Men and Machines in Language Translation*. <https://aclanthology.org/www.mt-archive.info/70/Kay-1980.pdf>
3. Memoq. *What is Machine translation*. <https://www.memoq.com/tools/what-is-machine-translation>
4. Smartcat help center. *Translate video, audio and transcripts*. <https://help.smartcat.com/translate-video/>
5. Smartcat. <https://en.wikipedia.org/wiki/Smartcat>
6. Translated. *New Research Confirms ModernMT Outperforms Leading MT and GenAI*. <https://translated.com/static-vs-adaptive-MT-vs-LLM>