Terminology management in computer-assisted translation (CAT) tools MemoQ as a case study

Tuka Anwar Subhi[1, a]

[1] Al Hikma University College, Baghdad, Iraq

[a] Corresponding author: tuka.anwar@hiuc.edu.iq

Abstract. Today, translation is the systematic thread that brings different civilizations and cultures together and brings them closer together. In the light of the considerable and rapid advances in technology, the responsibilities of the translator or of the translation companies increase on a daily basis. This led software developers to find tools to help the translator achieve those standards and enrich the translation market with new skills to handle them. Among these tools, perhaps most noteworthy is computer-assisted translation (CAT) Tools. The present paper sheds the light on one of the most important features in computer-assisted translation (CAT) Tools that is terminology management, furthermore, by taking MemoQ application as a case study, it shows a practical guide to create and use term base within this application.

Keywords: Terminology and translation.

INTRODUCTION

To talk about Terminology Management in computer-assisted translation (henceforth CAT) Tools and MemoQ, we need to know what are terminology management, MemoQ, and CAT Tools. "CAT Tools" is computer-assisted translation, which is something very different from machine translation. It provides several tools to facilitate and speed up the translation process. These tools can be downloadable software or recorded through the web browser. One of these programs is MemoQ. As for terminology management, it is a function of a premium CAT tool, as it stores the information that the translator wants to keep and may later want to use in special databases called "Term bases" and "Translation Memories", all of those are discussed below.

1. Terminology management

In addition to the need to have access to existing electronic dictionaries, terminology data collections, and terminology data banks, and the availability of Internet access for all kinds of terminology research and improvements. There has to be a mechanism and a tool to manage our "own" terminology [4].

Terminology means special or technical terms used with a particular technical application in diverse subjects of study, profession, etc. There is an enormous number of terminologies in this day and age into broad and crosscutting themes. In order to facilitate and improve the process of translation (computer-assisted translation), the translator needs to have his storage of terms and terminologies organized and managed in an instrumental system. Therefore, Terminology management has been set up.

According to Sarah Lewis, a technical writer, the definition of terminology management is "the systematized collection of concepts and terms with the purpose of unifying and standardizing technical documents. It can be carried out in a variety of formats including electronic databases or file glossaries" [2].

Under this definition, terminology management can be defined as a procedure, containing storage, authentication, and manipulation of terms. The terms here refer to special words. All translators need to highlight effective terminology management. At the same time, it is also important to ensure that terminology is properly managed by professionals and collaborators as well. It is then possible to stimulate compatibility and common symmetry with the source language. This contributes to strengthening the reading process. In addition to properly managing terminology, the translator should be able to use the updated terms applicable at the time the content is translated. At the same time, the translator should be able to store new terms in the database as well.

This will help the interpreter to ensure that they incorporate the correct terminology into the documents they produce. It must be ensured that terminology management is valid at all times because it can help save time and effort in the long term.

1.1 Terminology management systems

It is a program designed to facilitate complex translation tasks. It is an essential tool for implementing a global growth strategy in that it brings different team functions together to simplify translation and localization workflows across borders and time zones. For this reason, it is also known as the globalization management system.

Translation management systems are software that supports complex translation tasks. Most translation projects involve multiple teams, these teams may be located in different countries, and here TMS can combine the tasks of different teams.

1.1.1 Translation management system (TMS)

It is a program designed to facilitate complex translation tasks. It is an essential tool for implementing a global growth strategy in that it brings different team functions together to simplify translation and localization workflows across borders and time zones. For this reason, it is also known as the globalization management system.

The systems of translation management are software that supports complex translation tasks. The majority of translation projects involve more than one team. these teams may be located in different countries, and here TMS can combine the tasks of different teams.

1.1.2 The features that the translation management system offer

At the general level, although there are many providers on the market, there are basic features that are present in every translation management system:

Resource planning (including project, finance, and resource management), process management (workflow, collaboration, and content flow), language management (including translation, review tools, glossaries, and translation memory).

TMS is a comprehensive tool that combines separate translation tasks in one place. The ability to constantly create and update language resources, such as glossaries and translation memories, to track time and costs, create project reports, as well as automatically extract new content are essential functions. All of which makes TMS (translation management system) an indispensable tool for resource efficiency and reduction of overhead costs.

Other important features are being developed to new standards for translation management software, especially for cloud-based SaaS platforms:

Available language pairs, translation in context, live previews, automated updates for schedules and costs, security features, API integration, CMS connection, automated quality assessment.

These new standards raise the level of traditional providers and will soon be implemented on a larger scale as SaaS platforms find wide acceptance among translation and localization management teams.

1.1.3 The main benefits of the translation management system

Translation management systems complete tasks for the user that can be time-consuming. The state-of-the-art cloud-based TMS system takes a step forward, offering anytime, anywhere access, enhanced security, and enhanced collaboration tools. In addition, they largely eliminate version control issues during the transition of content between different users.

Traditional software solutions for managing translation projects require multiple software applications for different aspects of the translation workflow, slowing down processes, and creating more opportunities for errors and Omissions. TMS provides a comprehensive platform for the entire translation and localization workflow in one work environment.

1.1.4   Classification of terminology management systems

Terminology management systems are classified according to many typologies and subjects. Dr. Schmitz mentioned five classifications in his presentation that are complexity, entry structure, autonomy, software technology, and business aspects. First, the complexity obviously depends on the number of languages; it is progressive from monolingual to multilingual systems. Second, the entry structure of the system can be predefined, free defined, or hybrid (a predefined structure that can be modified). The third classification is "autonomy". There are autonomous systems that are identified or performed by a device capable of operating with no direct human monitoring, CAT tools component systems, and hybrid ones. The fourth classification concerns software technology as it can be systemized as a stand-alone, client-server, or browser-based system [4].

1.1.5   Concept-oriented and term autonomy

Organizing the entries of a terminology management system depends on three layers, which are concept, language, and term. A concept may have several terms (synonyms, abbreviations…, etc.) and all terminological data belongs to one concept including all terms in all languages, and all terms and administrative data should be stored in one terminological entry [4].

"Term autonomy guarantees that each permitted or prohibited term can be assigned the same number of features" [3].

Terms that belong to one concept (or to one terminological entry) should be organized and managed as autonomous blocks of data categories without any preference for a specific term.

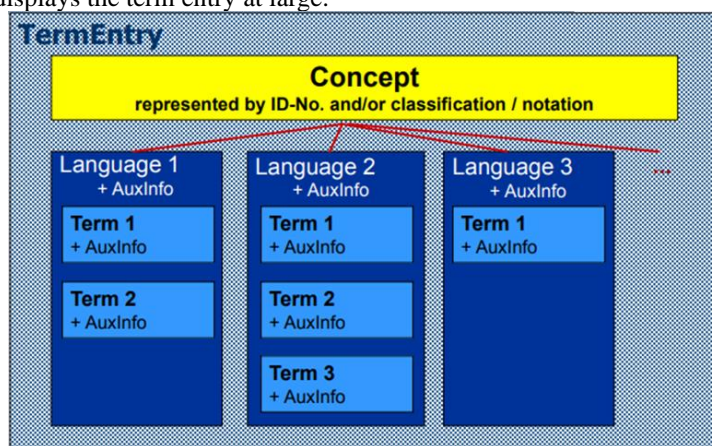The figure below displays the term entry at large.

Figure1, Term Entry

2.   TERMINOLOGY MANAGEMENT IN CAT TOOLS

A human translator has the ability to maintain consistent use of company-approved terminology while using CAT tools; he/she can store a downloadable glossary (or a term base). This is particularly essential to the proper translation of technical, scientific, and legal texts, which require a high degree of consistency and accuracy. Most glossaries start in Microsoft Excel and must be converted or saved as a TBX file to be used by a CAT tool.

Terminology Management is the process of identifying, storing, and managing company, customer, or product-specific terminology that needs to be translated in a specific way. Terminology or terms are stored in term bases, which can be accessed in a CAT tool. A term base is a searchable database or glossary containing multilingual terms, references notes, and rules about their usage. Translation Memories (TMs) store whole segments while term bases contain single words or expressions.

2.1 Terminology management in MemoQ

Through the MemoQ program, it is possible to create a terminology base, translation memory, and consistency of translation of terms and to consolidate the efforts of translators involved in translating a project (Jenie, 2019).

MemoQ provides advanced tools to assist with the translation and management of control terminology. MemoQ tools are flexible and enhance the process of creating a rich repository of documents and databases, which is very important to large organizations.

It is worth mentioning the advanced features in MemoQ that facilitate managing terminological information.

- Import terms to program database:

The translator can import a database of previous terms you may have or may the client provide. If there is a term that we have added more than one synonym for, we can, by importing the terms in the program, import and even customize a list of previous terms that we have, for example, include the terms as the basic meaning, a synonym, a note that explains the meaning of the word, or even a forbidden translation. All terms will be automatically organized by MemoQ.

- Advance translation:

You might happen to be working on the translation of one of the documents that bear a great resemblance to the many projects you have worked on before. In other words, many repeated sentences will be kept in the translation memory. Through the pre-translation option, you can request MemoQ to translate all duplicate sentences or sentences similar to the translation memory content of the current document directly. The program tells us that we want to translate all sentences that have a good resemblance to "good TM or corpus," which is usually more than 60% similar.

- Predictive typing:

During the translation and writing process in the language box, the goal is to make the MemoQ program predict the word you want to write. This feature is particularly useful when working on documents that contain relatively long or complex words.

- Auto pick:

Another very useful feature in the case of working on documents with numbers, equations, codes, links, etc., is automatic selection. Through which MemoQ allows you to copy all the details except the vocabulary in the original sentence to the target sentence quickly and with a single click.

- MemoQ web search:

websites that are relevant to the project can be used, such as dictionaries, encyclopedias, search engines, and other sources that you may need while working. The software shows you these sources directly at any time you want. You can activate the online search feature by pressing the (Ctrl + F3) buttons, but you first have to set up the sites you want and provide MemoQ with their own links.

2.1.1 Term base

Terminology is the greatest burden in the translation process, especially in the field of specialized translation, which requires extreme accuracy in the choice of terms to produce a correct translation of the text, such as medical, legal, economic, and other fields with their terminology and without which translation is not possible. Which saves a lot of searching effort and unites them into joint translation projects. A glossary or glossary of terms, keeping them consistent in the translation process, is one of the most important characteristics of translators.

Term base is more like a glossary, a database containing words or phrases that are often bilingual or multilingual.

2.1.2 The working mechanism of term bases in MemoQ

Practically, the term base is a group of terms that are collected and categorized by a human translator, containing single words, phrases, or expressions related to certain subjects. These terms can be bilingual or multilingual.

MemoQ's term bases are multilingual: a term base entry is not restricted to one source and one target language. A term can occur in any number of languages within an entry.

Term base is an integrated feature in MemoQ. In this program, the user is free to import existing term bases or build his own ones. The user can also add or update terms as he translates and specifies terms as forbidden. Terms are added in a quite simple way, any terminology can be sent to the term base by marking a word or a combination of words and saving them by clicking on the option "quick add term", then they will appear in the results list directly and also showed in a field below the list.

Terms of the user term bases are marked in blue in the text and shown with their translation in a list of translation results, the forbidden terms also appear in the results list but they will be marked in black. While writing the initials, the program will show you a list of words with the very first initials, you can choose the appropriate word or you can continue writing without choosing any word, and the writing rate or style will not be affected (the screenshot below displays these steps). This feature makes the process easier and smoother.

Another characteristic of MemoQ is that the program enables the user to collect his own terminology and combine several bilingual term bases into multilingual ones.

MemoQ also provides an important feature for excel files when there is already a list of terms. The files will be organized in columns that are assigned to the different fields in the term base.

2.1.3 The benefits of using a term base

- Increases consistency:
  Consistency is essential while working on any project at any organization, especially the ones that have several collaborators involved. Therefore, the translator should build a suitable term base that enables him to keep the core message of the project consistent.
- Improve translation quality:
  In a term base, the translator can manage terminology and define terms according to specific criteria, such as defining forbidden terms. The forbidden terms (unwanted words or expressions), for instance, will not be used or repeated by the translators.
- Speeds up translation:
  Using a term base makes the translation process simple and straightforward. In CAT tools programs, including MemoQ, there is auto-correction or quick suggestions that help the translator replace words without suffering for the sake of finding the proper word, this feature helps the translator to save time and effort.

2.1.4 Building a term base with MemoQ

a) building a term base from a specific project: Term bases can be built before starting a project by setting a specific term base for the project in question, then when working on translation specific words, expressions, or entire sentences can be added to the term base as shown in figure 2 below.
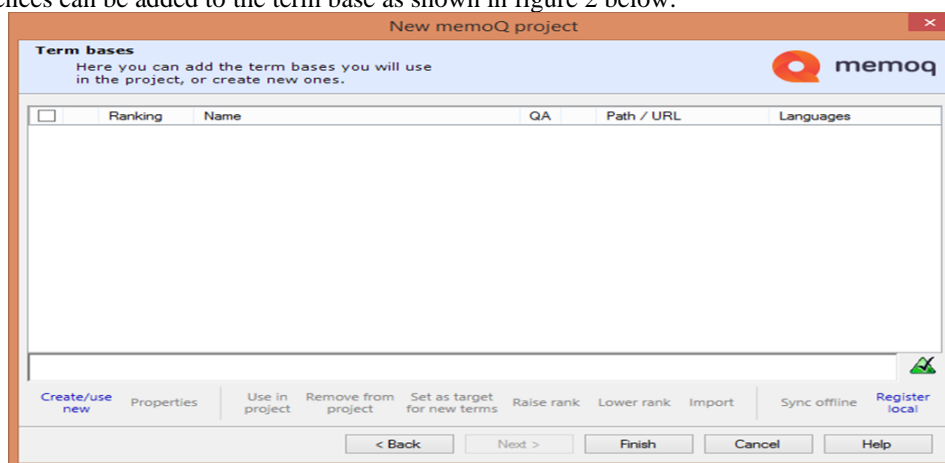

Figure2, creating a term base in a new project in MemoQ

Then, the term base can be named and the language pair can be set, while other information including the project's domain, subject, and the client name, will be filled automatically as shown in figure 3 below.
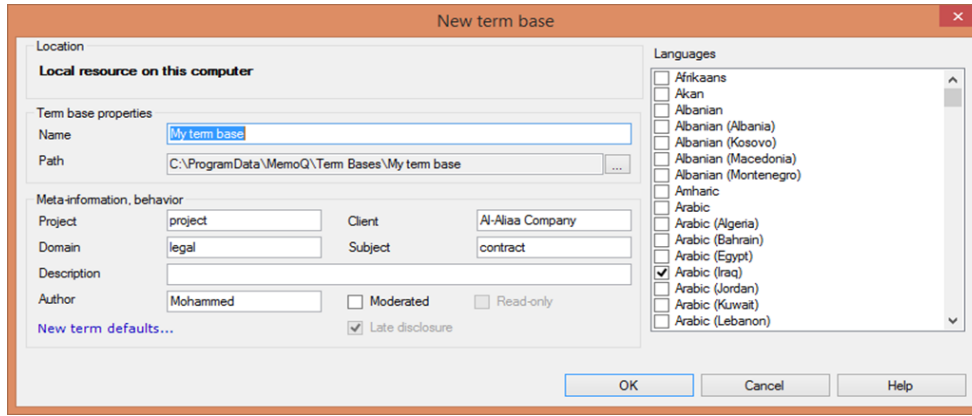
Figure3, setting properties for the term base

To send terminology to the term base, while working on a translation project, you only need to mark the terms, which can be single words or word combinations, and save them as shown in figure 4 below.
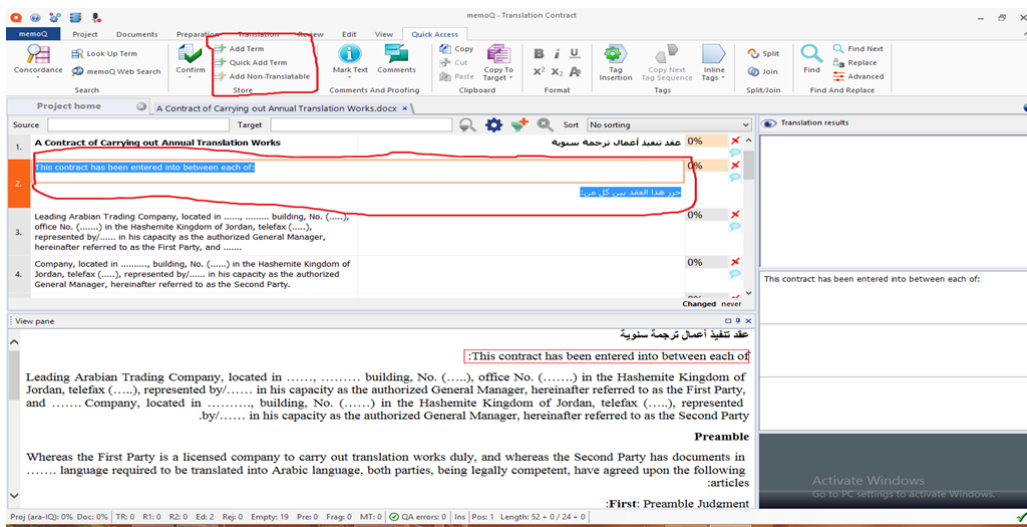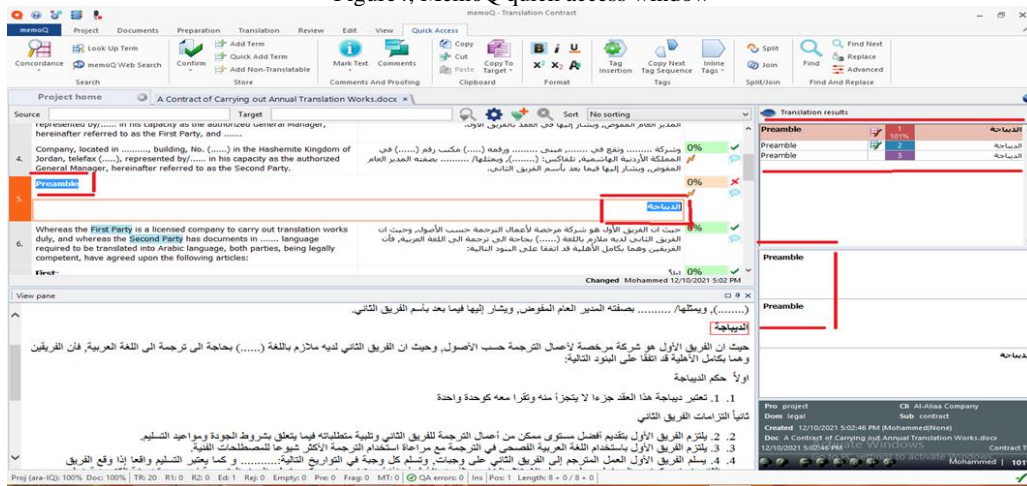

Figure4, MemoQ quick access window

Figure5, MemoQ quick access window

Terminologies can also be modified within the term base, for example, to specify whether the term is matched as usage, grammar, or definition. In addition, matching percentage can be set whether as exact or fuzzy match.
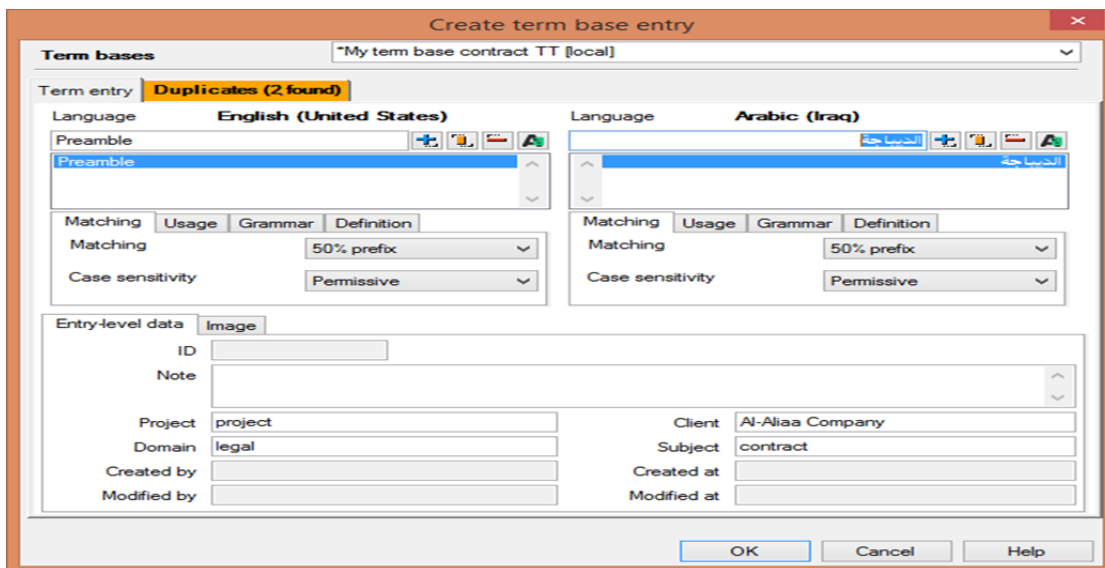
Figure 6, term entry window

The chosen terminologies will appear in the results list directly and can also be viewed in the field below the list.
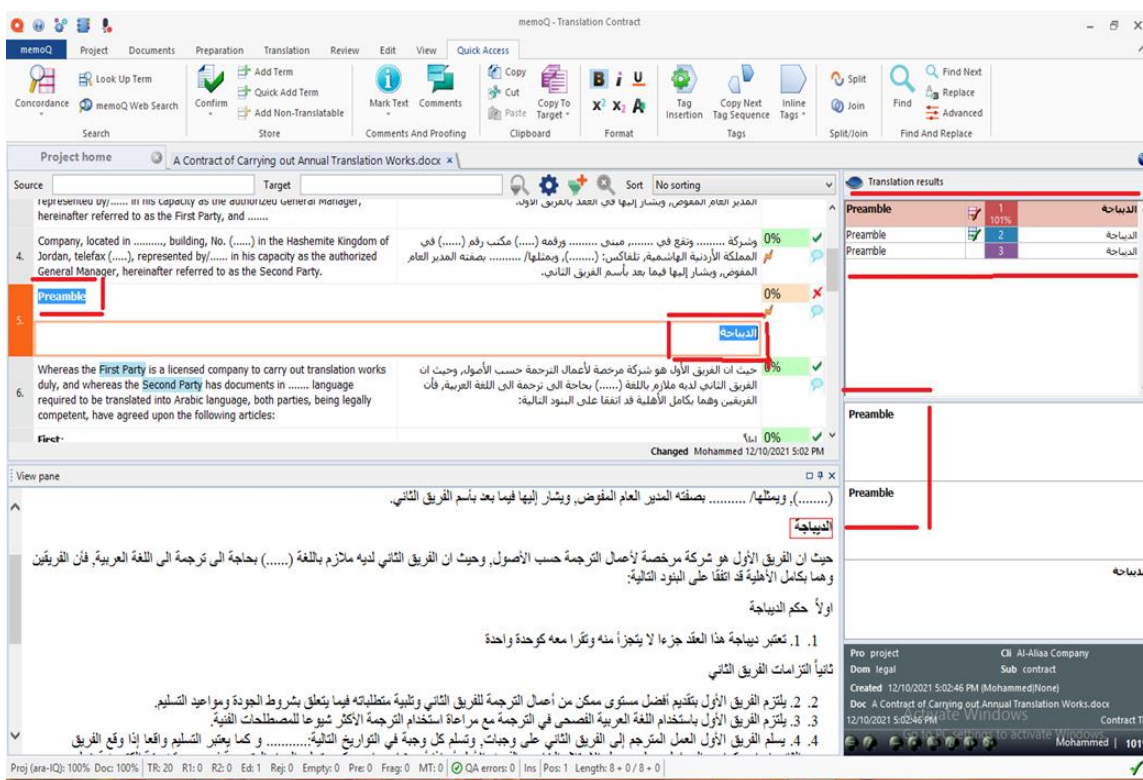
Figure7, MemoQ quick access window

When the terminologies are saved in term base, they can be reused in further projects or within the same project. The application will automatically suggest the translation of the terminology in question that has been previously added to the term base. As shown in figure 8 below.
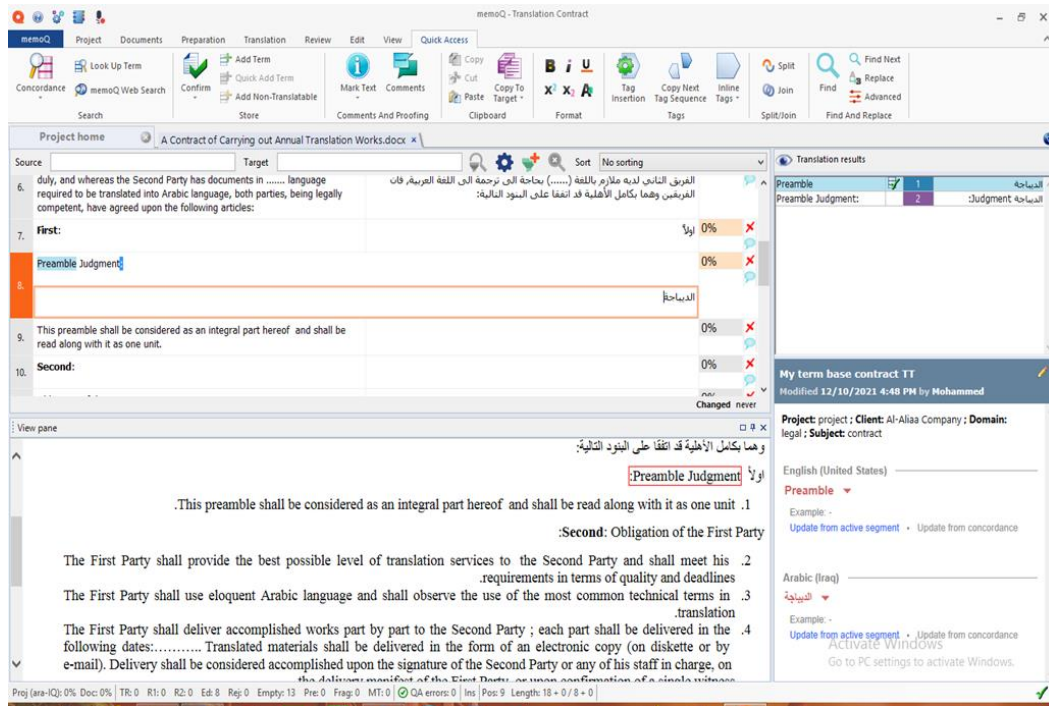
Figure8, MemoQ quick access window with results list on the right

Furthermore, terms can be extracted from the same document, to ensure full consistency in terminology by identifying and translating frequent expressions before the translation project starts as shown in figures 9 and 10 below.
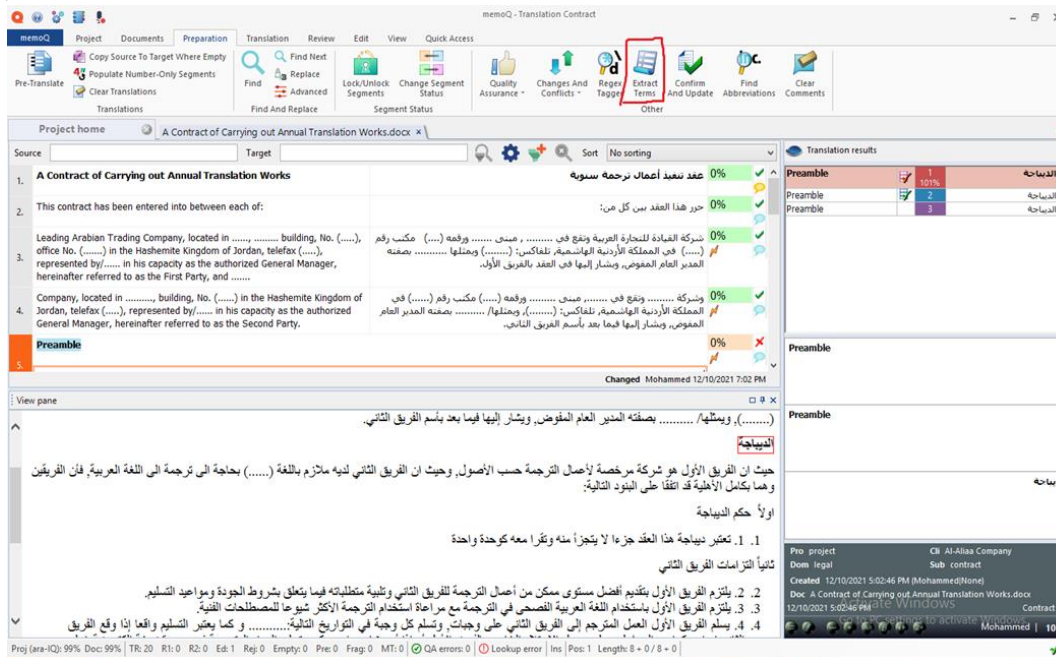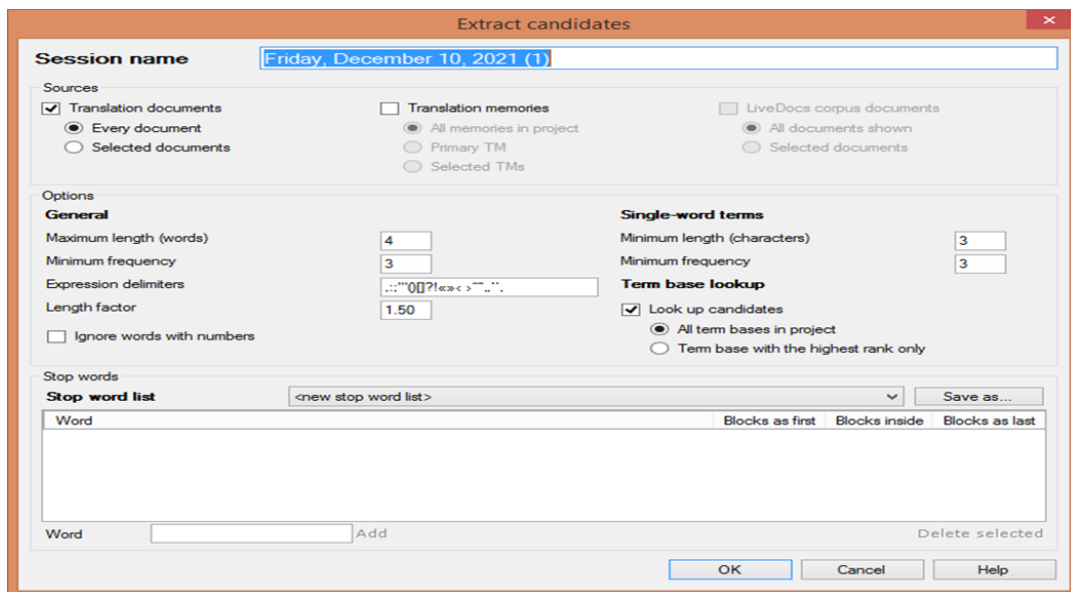

Figure 9

Figure10, candidates extraction window

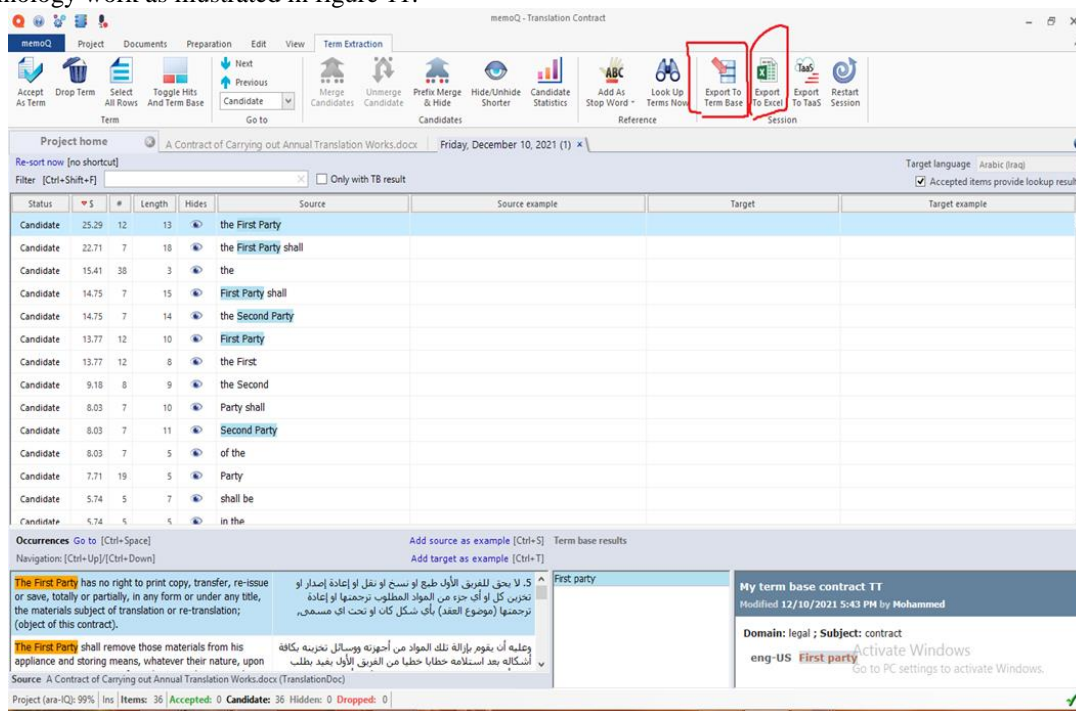Then, the accepted terms t can be exported to the term base or to an excel sheet for validation or additional terminology work as illustrated in figure 11.

Figure11, term extraction window

In addition to the built-in MemoQ term base where terminologies can be collected, MemoQ is also connected with various online dictionaries, thesauruses, and search engines. Preferred online dictionaries can be set from the setting and the result of the search will be shown from all the chosen online dictionaries.
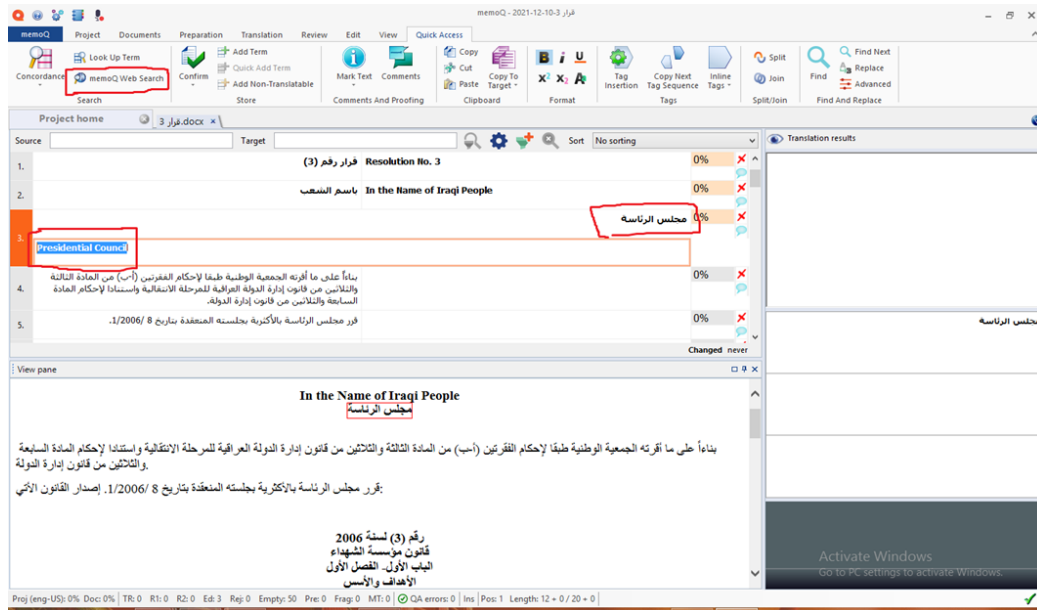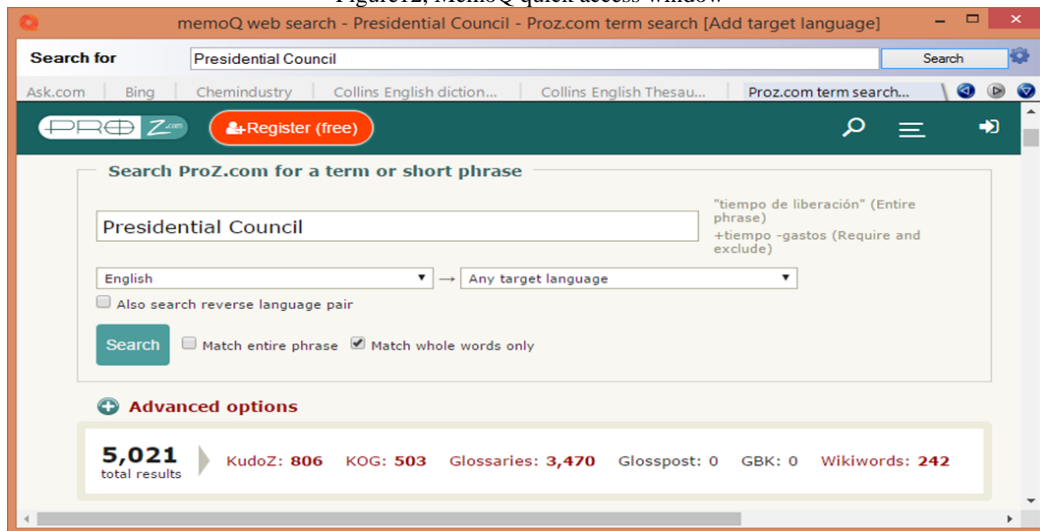
Figure12, MemoQ quick access window

Figure12, MemoQ built in web search

When the project is finished, the term base will be automatically stored and can easily be found on the interface of the application, where it can be edited, removed, or used with further projects as shown in figures 13 and 14.
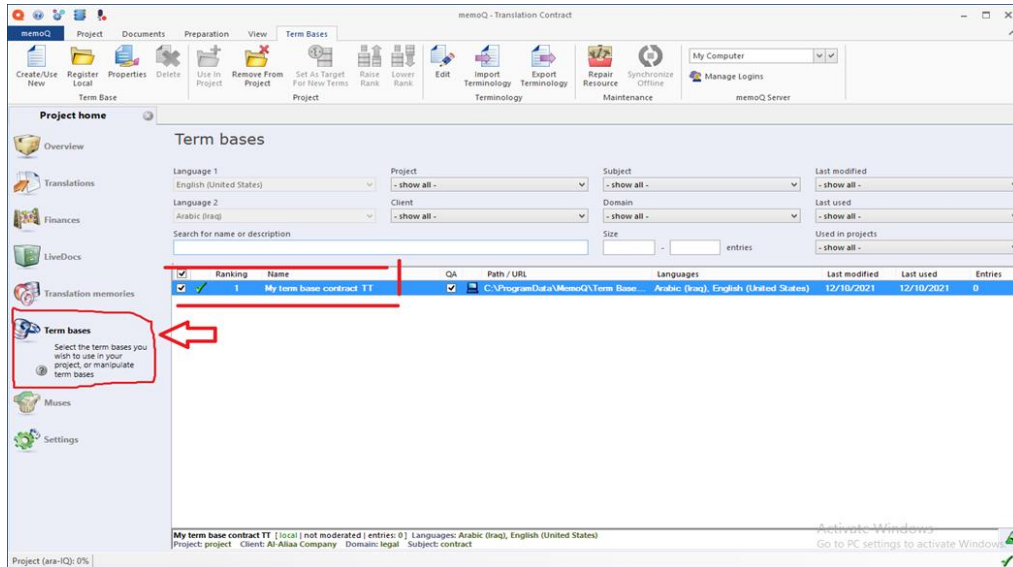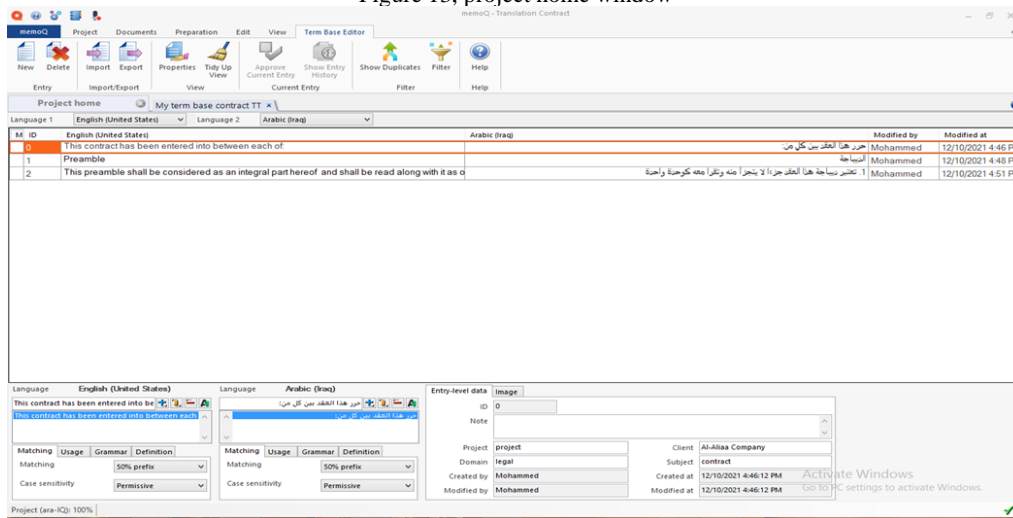
Figure 13, project home window

Figure 14, term base editor

b) creating a term base from an excel sheet: term bases can be created and used from an excel sheet by importing an excel sheet containing terminologies in a language pair, the language pair can be specified in the excel sheet or in the application itself. For example, a legal glossary in the form of an excel sheet can be added to the term base as shown in figures 15, 16, and 17 below.
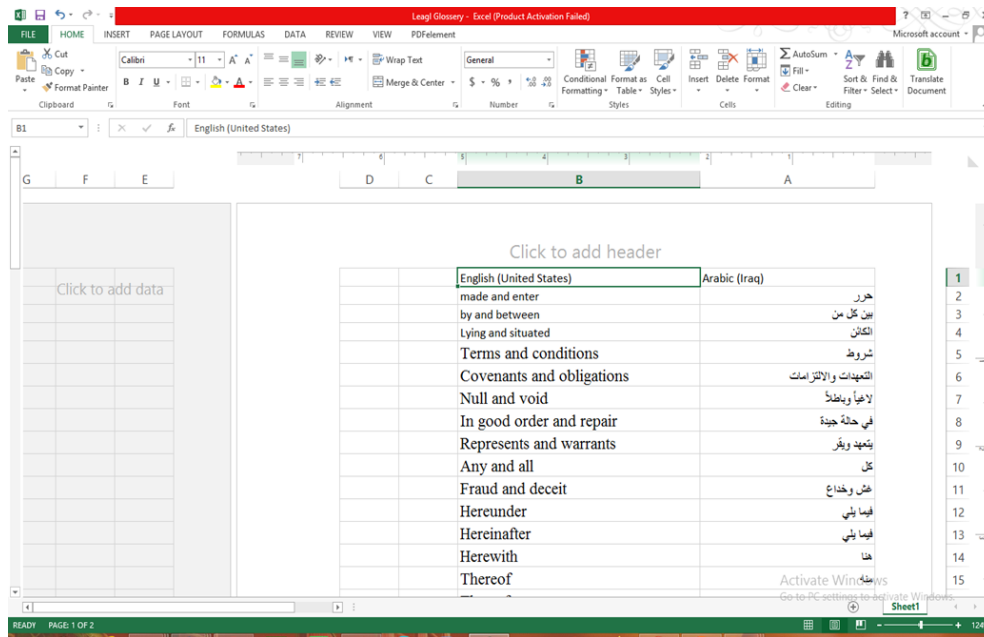
Figure 15, sample of excel sheet

Then, the glossary can be added in an old term base, or it can be added to a new created term base.
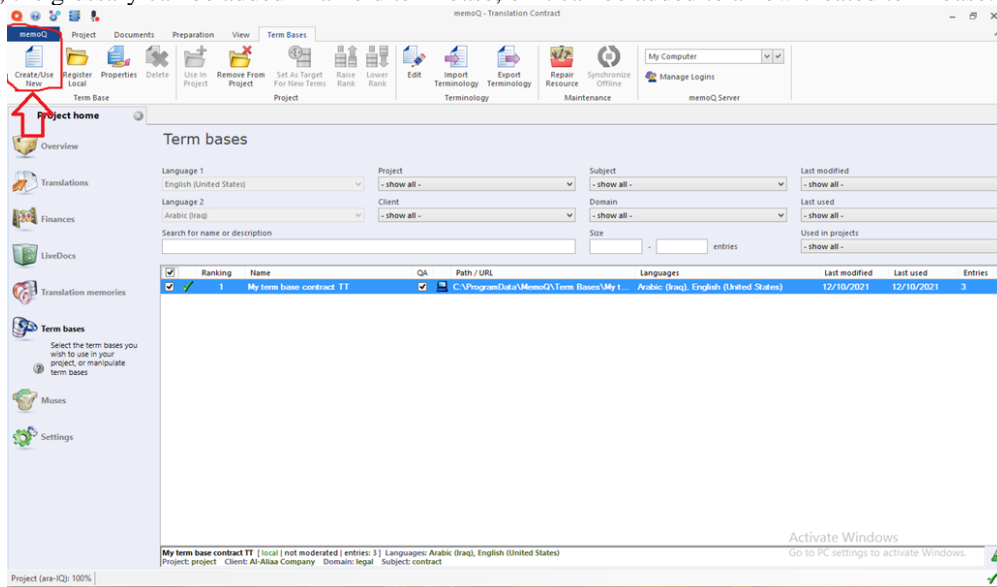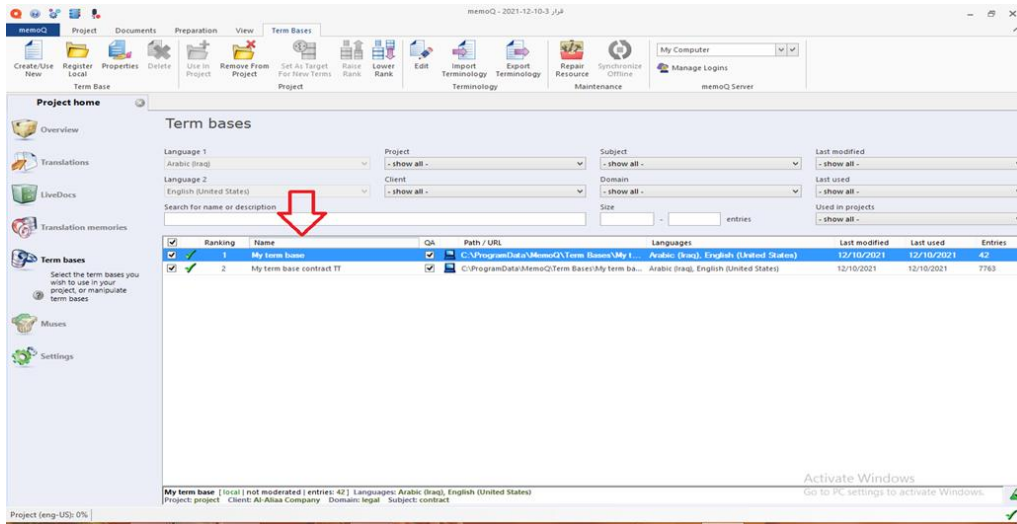

Figure16, term bases window

Figure17, term bases window

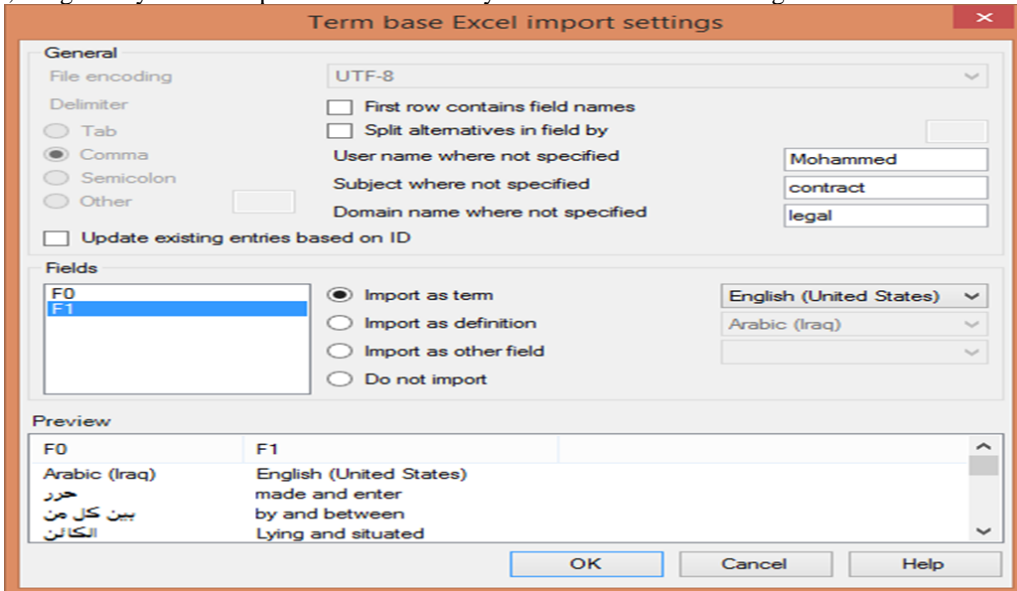Then, the glossary can be imported from the newly created term base. See Figures 18 and 19 below.

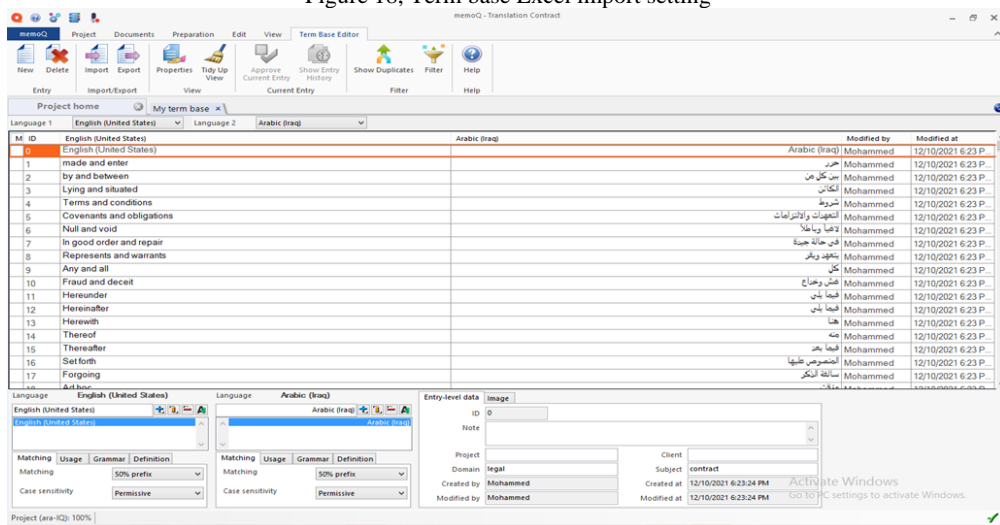Figure 18, Term base Excel import setting

Figure 19, Term Base Editor

2.1.5   The importance of terminology management in translation:

The translator must adhere to certain basic rules to preserve the two types of translations that it performs. This will significantly improve the productivity of the translator's work. One of the most important key rules is adherence to terminology management.

2.1.6   The advantages of terminology management

Having gained a thorough comprehension of terminology management, we can delve into its significance. Below are key points demonstrating the importance of this field, offering insight into why translators should pay closer attention to terminology management.

- It can ensure consistency and symmetry:
  Terminology management is one of the most important factors for achieving benefits in translation and its role is to assist translators to ensure consistency in the work to be done.
  Ensuring consistency in translated terms is crucial for achieving accurate results consistently. Updating the terminology database regularly is essential to maintain this consistency, particularly when integrating upgraded terms. This practice guarantees coherence in translation outcomes.
- It can save both time and money:
  With proper terminology management, one will be able to provide people with the unique help of saving time and money as well. Translators need to get terminology from their available database. Besides, the spent time completing the translation subject could be reduced to as little time as possible.
  The above is equivalent to money, so saving time in translation helps the translator save money as well. This also contributes to reducing the pressure on the translator as well as the client. Finally, proper, and smooth translation can be produced in less time and at less cost.
- It also improves quality in translation:
  Another advantage linked to terminology management is quality control. It is imperative for translators to continuously strive for improved quality, as their success hinges on the accuracy of their translations. With a solid foundation in terminology management, it becomes feasible to consistently employ the correct terms in their appropriate contexts, thus enhancing the overall quality of translations.
- Experience to offer a better field in the audit process:
  After the translation process has been completed, it is important to check the document. Extensive and good experience in auditing can be obtained with the help of the terminology management process. In this case, the errors that can be found are few and even rare. In addition to the above, maintaining a minimal translation memory becomes feasible. Additionally, keeping translation memory low leads to superior and quicker results compared to the final review stage.
- Working in different languages is the best option:
  Translators and translation offices work in different languages. With such a large number of languages handled, the translator will often recognize specialized requirements. Every language has new terms and conditions. If a proper terminology management department is available, there is, therefore, no need for translators to mention these specific terms and conditions. Mentioning these terms and conditions can be a burden for the translator. This burden can also be reduced by supporting and assisting terminology management. This thus preserves the coherence of the texts provided to the client.

The subject of terminology management must be taken into account continuously. Then you will be able to access the best benefits of the translation work you do.

3. TRANSLATION SEGMENT

Segmentation in translation involves dividing the source text into smaller translation units. These units adhere to specific segmentation rules, serving as the foundation for building and refining the translation memory tailored to the selected language pair. These rules represent an advancement in translation automation, enabling systems to automatically recognize and apply them during the translation process.

The default segmentation rules compare with the details of each supported language and can be modified. Things imported with a terrible segmentation, for example, inefficiently designed log documents or using the wrong allocation of segmentation, can affect the attributes of the coordinates of the translation memory [8].

3.1 Focused aspects of segmentation rules in translation

When a translation document is added to a task, it is split during the import interaction (see figure 20). Partitioning is a cycle of dividing content in a translation history into translation sections or units. The work of the interpreter then translates each unit, preparing a translation pair consisting of the first part, called the source section, and its translation is called the thematic part that can be placed in the translation memory. Of course, the boundaries of the part are confirmation signs. In general, the whole point indicates the end of the sentence, and a sentence is usually an important unit with a possible translation. However, there are exceptions. Might happen, for example, that's the whole point not followed by another sentence, for example, due to the ordinal number in certain accents.

Computer-assisted translation plays a vital role in implementing segmentation rules. The introduction of computer-assisted translation tools has provided interpreters, commentators, and external contractors with access to software that substantially reduces the time spent on translation-related tasks through computational means. Segmentation represents a fundamental step in the processing of source content before it is engaged with for translation. Through segmentation, the source content is divided into translation units known as fragments, which can include sections, bullet points, headings, and other elements.

Specific segmentation rules naturally generate segments, aiding in the creation and customization of translation memory for specific tasks or clients. This division serves as a fundamental structure for future use, leveraging existing interpreted material. By utilizing these segments, both the interpreter and analyst find their tasks simplified. Focusing on and recalling the precise wording of a section within a large-scale project can still pose a challenge. However, segmentation and the capabilities of translation memory help mitigate these translation issues by preserving recently deciphered segments. Through segmentation and translation memory alignment with decoded sections, the computer-aided translation (CAT) tool seamlessly integrates the current translation into the exact context with the same material or replaces it seamlessly as needed.

Segmentation rules can be adapted for each task. When selecting a task, the CTI tools focus on selection, where the client will choose whether to use pre-arranged segmentation rules or highlight specific rules. The rationale behind segmentation groups can vary. The totality of the highlighted rules will be put away in the SRX register. Similarly, special cases of rules can be described.

Translation rules can be modified. There are also some definitions such as:

- The full point determines the end of each part
- The punctuation mark marks the end of each syllable
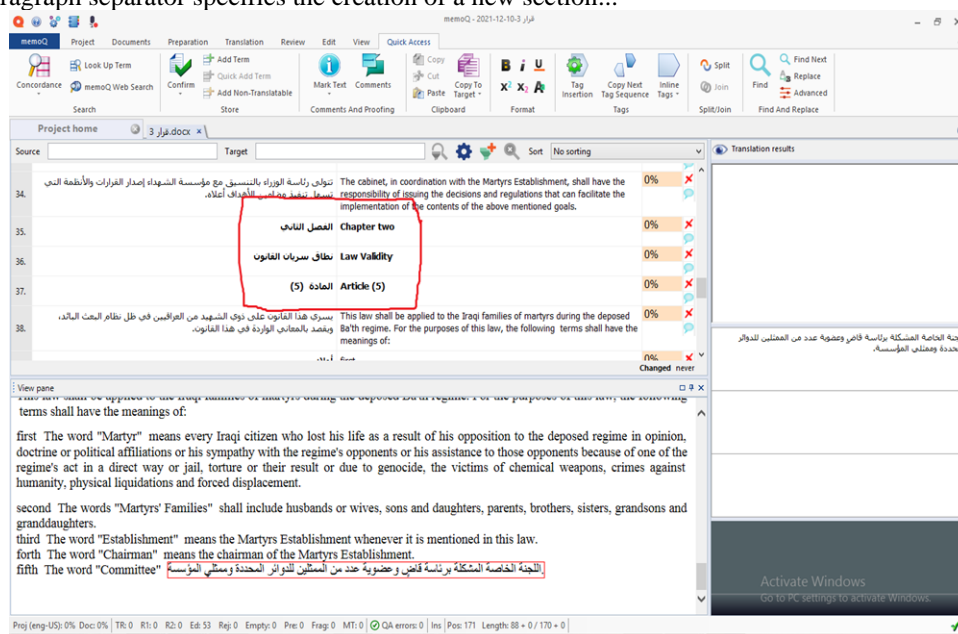- Paragraph separator specifies the creation of a new section...

Figure 20, quick access window with translation segments

4. TRANSLATION MEMORY (TM)

Translation memory (henceforth TM), also known as translation databases, is a database for sentence pairs and constitutes the heart of CAT Tools. It is a collection of entries where the source text is associated with its corresponding translation in one or more target languages, translation memory is usually used in translation tools: the text program is divided into segments, which can be blocks, pieces, sentences, and even phrases when the translator opens a slide, the application searches for the database of the source.

Translation can perform much faster: unnecessary and mostly rewritten translations are avoided for text in need of change; in addition, translation memory allows better quality control by providing translation to already approved candidates with the correct terminology.

Translation memory is a powerful technique that can help reduce the cost of localization, however, the use of translation memory needs to weigh all factors that are taken into account because the application program is effective in translating the text, a high degree of repetition.

Translation memory (TM) reuses previous translations to improve translator productivity, works by splitting text into short sections and storing the translation in a database and when repeating a section, the translator can reuse a suitable previous translation or sometimes paste a translation to a similar segment with minor modifications.

4.1 The point of using translation memory

Cost-saving and speed increase
Legal contracts, software interfaces, technical documents, medical posters, and product catalogs contain frequent texts according to the study of Memsource when translating these types of texts the translation memory can reduce costs and increase speed by 15-90%.

Any team member can reuse translations from other members, saving time and improving translation consistency.

More than 90% of translation companies use TM and CAT tools, according to a 2016 survey to save costs, and translation memory is the industry's core technology.

4.2 The proprietary of translation memory
Translation memory contains actual translation, so it is treated as intellectual property by law in most countries that limit the translator's ability to download the database and reuse it for another client.

4.3 The use of translation memory
Stored locally or in the cloud, the subtitle memory can be used collaboratively and easily transferred in a compressed file format. It is an ever-growing database that serves as a source of transition for future translations and is, therefore, an important economic asset. Computer-aided translation tools (CAT tools) and TMS (translation management systems) with translation memory functions enable the creation of translation memory easily and smoothly. Translation memory can be created and classified according to some of the following criteria:

For every language pair, for every subject, for every industry, for every customer, for every project, for every user or owner.

Accessibility and usability can also be managed. This concerns user groups but also potentially confidential or proprietary information contained in the source.

For example, project-based translation memory can be created that may contain special content or intelligence based on previous industry-related translation memories (see figure 21). At the same time, the newly added translation content can be kept subject to the project's translation memory only.

The method of creating the translation memory can also be managed. Since TM is chip-based, applying specific segmentation rules that work best for the translation project at hand would be a useful strategy.
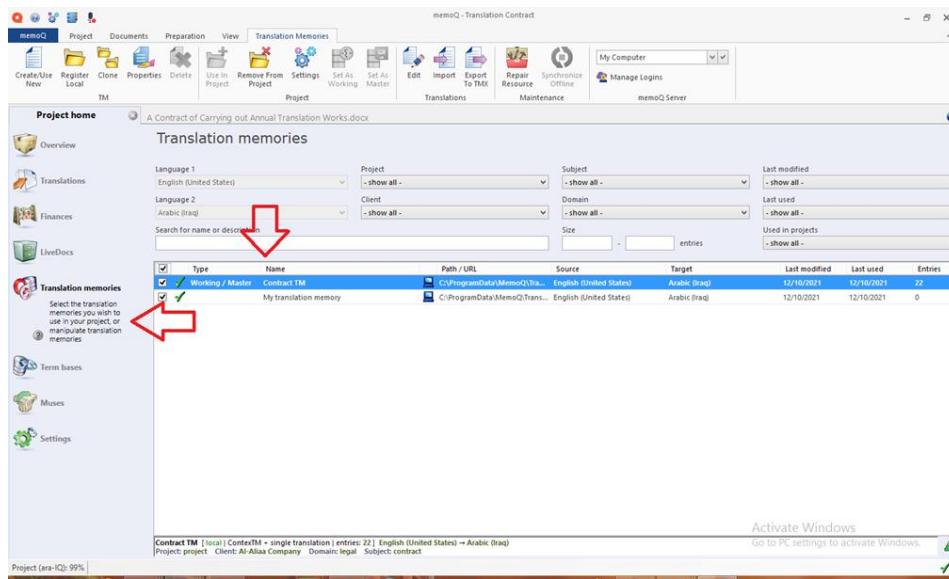
Figure 21, translation memories window

The TM will show suggestions for translation for identical or similar sentences in the results list in red. The similarity is shown by the percentages the differences between the sentences are marked with colors as shown in figures 22 and 23 below.
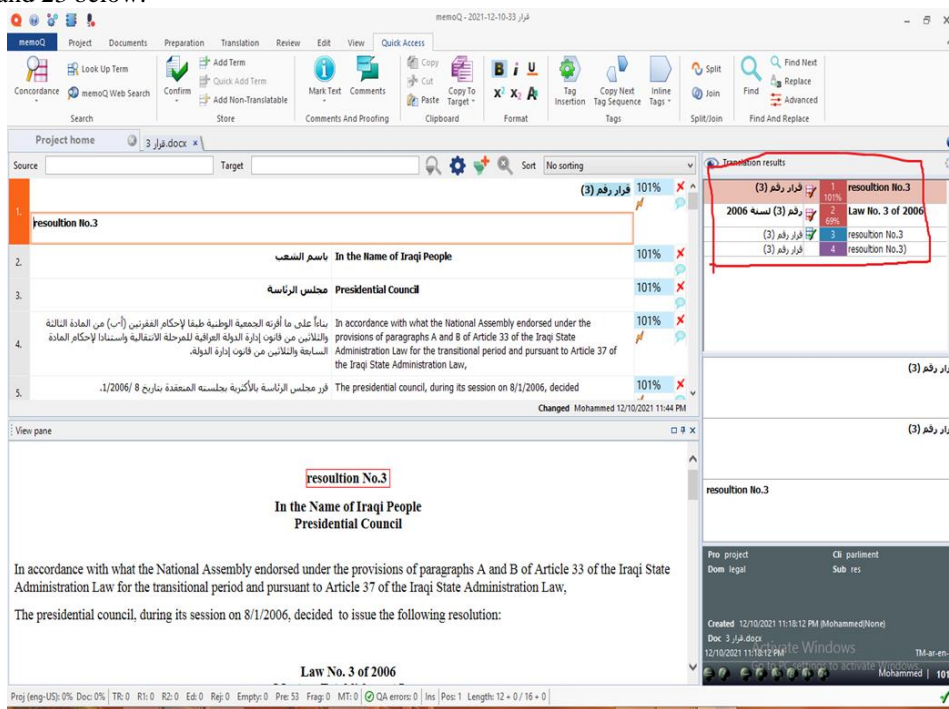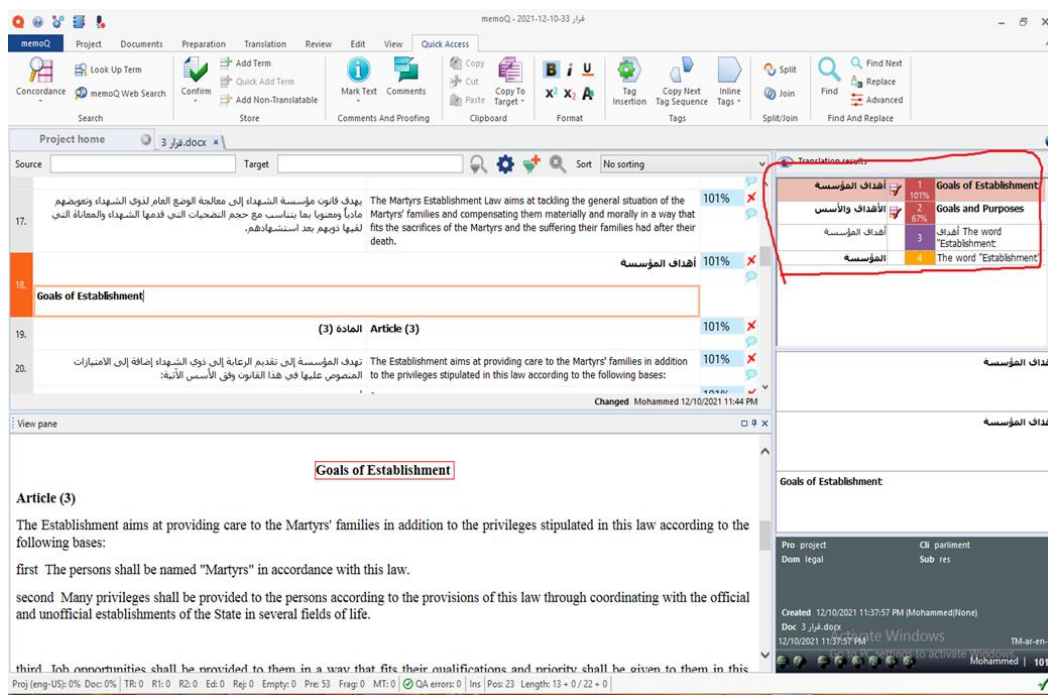

Figure 22, quick access window

Figure 23, quick access window

The TM also looks up words and phrases in other sentences through its automatic concordance search shown in yellow in the results list.

Concordance search can be run manually and insert useful parts from the concordance window as shown in figure 24 below.
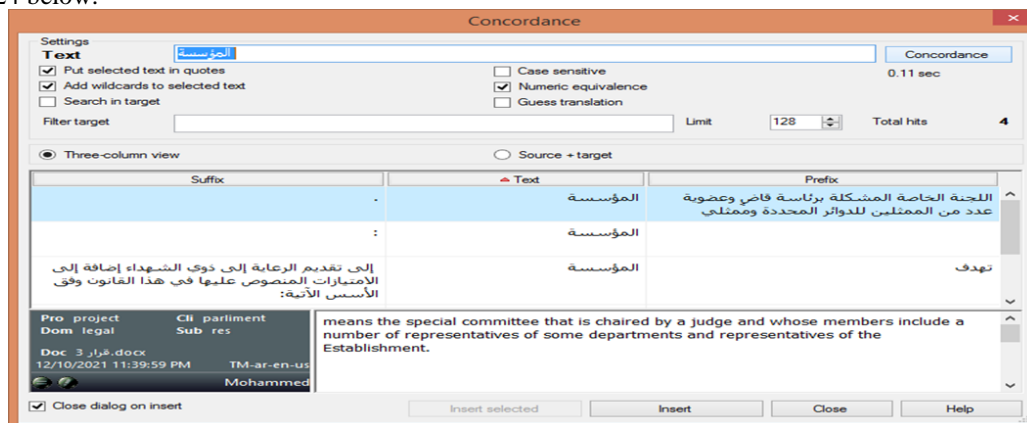

Figure 23, quick access window

4.5 Confirmation of the quality of the translation memory

Ensuring the quality of translation memory is just as important as creating and updating it regularly. Especially in cases where multiple users contribute to translation memory, reviewing and debugging the database for the best use cases and applying terminology according to context is essential.

This stems from the fact that in a variety of languages there can be no term that has different meanings depending on the context, and therefore require a different terminology in the translation. It is also important to keep it as clean and topical as possible.

Usually, be deleted entries from the translation memory are only possible for administrators and project managers. Best practices suggest asset building to achieve consistency and natural outcomes. To achieve this, Guidelines, terminology rules and glossaries shared with the creator and user of translation memory can go a long way.

Translation memory will be useful only if its quality is maintained at a constantly high level. This will improve matching results in future translations and provide accurate context for translators working on any project

CONCLUSION

After all, of the above, it becomes clear to us the importance of using CAT tools in general and using terminology management.

Relying solely on Electronic Technologies is not a cure-all solution for educational challenges. While their aim is to expedite the translation process, mitigate various issues that arise, and reduce the time required, they cannot replace the expertise and language proficiency of a human translator. Even with advancements in artificial intelligence, personal involvement and professional experience remain crucial factors that cannot be substituted. Electronic systems, though valuable and indispensable in certain scenarios, cannot compensate for a translator's lack of familiarity with these tools, which can be a significant drawback. However, if translators possess a thorough understanding of how to leverage these resources effectively, they can significantly enhance the translation process. Therefore, translators require software that integrates both translation capabilities and tools for proofreading and refining drafts comprehensively. Editorial programs have been specifically designed to aid professionals, such as lawyers and office personnel, in enhancing the quality and clarity of written communications. Translators can utilize these programs to craft English content in an international style that is clear, concise, and engaging.

REFERENCES

1. Debunking common myths about CAT tools (2017, December 15). Debunking common myths about CAT tools.. https://atril.com/2017/12/15/debunking-common-myths-cat-tools/
2. Lewis, S. (2018, October). Terminology management. Definition. https://whatis.techtarget.com/definition/terminology-management/
3. Mackenthun, A.(2013, may 03). Terminology Management Systems https://www.parson-europe.com/en/knowledge-base/terminology-management-systems
4. Schmitz, K. (2011, Juli), Tools for Managing Terminology http://www.termnet.org/downloads/english/events/tss2011/9_TSS2011_KDS-TermTools.pdf
5. Gabriel, D. (2018, April 18). Introduction to Computer-Assisted Translation. GMD linguistics. https://gmdlinguistics.com/introduction-computer-assisted-translation/
6. Jenie, G. (2019, October 9). 16 online translation tools recommended by translators.
7. Gengo. https://gengo.com/community/15-online-translation-tools-translators/
8. Will, K. (2021, August 01). Segment. Investopedia. https://www.investopedia.com/terms/s/segment.asp/