pairaphrase

Edit and improve translations

pairaphrase

Actions

1

2

3

4

Use the same translation technologies that professional translators use

BEST PRACTICES FOR AI-POWERED GLOSSARY DEVELOPMENT

Updated For 2025

BEST PRACTICES FOR AI-POWERED GLOSSARY DEVELOPMENT

Creating a glossary for machine translation using AI versus a glossary for human translation involves different considerations and processes due to the unique requirements and capabilities of software compared to human translators. Both have their strengths and weaknesses. This guide was designed and developed to help machine translation users take advantage of AI-powered translation.

WHAT IS A TRANSLATION GLOSSARY?

A translation glossary is a bilingual database file created by an organization that compiles key terms and their consistent translations to ensure accuracy and uniformity in multilingual texts. It serves as a reference tool for translators, providing pre-approved translations for specialized vocabulary.

AI-powered glossaries are used by the translation engine to enforce specific translations automatically, by inserting terms directly into the translation. This ensures consistency and uniformity and is particularly helpful for industry-specific jargon, brand names, technical terms, and frequently used phrases.

Glossaries help maintain consistency across different documents and projects, they reduce the risk of errors, and ensure that translations adhere to the desired terminology standards. Glossaries are particularly valuable in fields requiring precise language, such as legal, medical, and technical domains, and can be used both in human and AI translation processes to enhance overall translation quality. This guide focuses on AI translation using glossaries.

HOW ARE GLOSSARIES HELPFUL FOR TRANSLATORS?

Glossaries help users produce more accurate translations and save time. Here's an example. Let's say you translate the following sentence into German (glossary terms are underlined):

"We at ACME Inc.'s Sales department hope that you will find our offer for ACME 2x0.5 inch Robo Widget Connectors acceptable and look forward to your purchase order."

Without a glossary, the translation engine gives you:

"Wir, die Verkaufsabteilung von ACME Inc., hoffen, dass Sie unser Angebot für ACME 2x0.5 inch Robo Widget Connectors akzeptabel finden und freuen uns auf Ihre Bestellung."

This is a correct translation, but it's not accurate, since your company uses different terms in German:

Source term	Machine translation	Your company's term
ACME Inc.	ACME Inc.	ACME GmbH
Sales department	Verkaufsabteilung	Vertriebsabteilung
ACME 2x0.5 inch Robo Widget Connectors	ACME 2x0.5 inch Robo Widget Connectors	ACME 5x1,25 cm Robo-Dings-bumsverbinder
offer	Angebot	Kostenvoranschlag
purchase order	Bestellung	Auftrag

If you have repetitive important words or phrases, or you find that certain terms require correction over and over again, and you're left with a lot of repetitive manual editing. This is where a glossary can help.

BEST PRACTICES FOR USING TRANSLATION GLOSSARIES

1　**Try and create your glossary before you translate a document.** This can mean skimming some text manually for potential glossary terms or using a keyword extraction tool. In any case, the glossary must be uploaded and synced before it can be used for AI translation. You can add terms to an existing glossary as you translate, but they will only be used in future translations.

2　**Most of your glossary terms will be nouns or set phrases.** A typical use case is company-specific terminology, like brand names, product names or acronyms. So let's say you have a product named "ACME Super Fun Toy". That belongs in your glossary. However, the word "toy" by itself does not belong in the glossary, because it can appear in too many different combinations and be either a verb or noun. Furthermore, machine translation will usually recognize the meaning of "toy" in context, so you do not need to standardize this term.

3　**Keep your glossary lean and include only words that you want to control.** If multiple target terms are possible, do not add a glossary entry. For example, "department" can be both a division within a company as well as a ministry, so it would make no sense to add "department" alone to your glossary. Instead, add unambiguous combinations, such as "Sales department, Vetriebsabteilung" or "Department of State, US-Außenministerium".

4　**Think of your Glossaries as fluid assets.** Plan to add new terms regularly and possibly retire existing terms as well. Glossary development is not a one-time exercise.
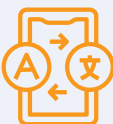
5　**Some translation engines can only recognize the exact shape of a word.** This means the translation engine will not recognize inflected or upper/lower case word shapes. So for example, if your glossary contains "Sales department", it will not recognize "Sales departments" or "sales department". If special word shapes occur frequently in your texts, add separate glossary entries.

6　**Glossaries do not work as well with languages where words change their shape a lot (inflection), such as Slavic and Semitic languages or Turkish.** The reason is that the translation engine only inserts exact glossary entries without grammatical adjustments.

7　**While glossaries do work for less-inflected languages, such as German or French, you still have to check the output to make sure any nouns and modifiers are inflected properly.** For example, "the size of the State Department" in German is "die Größe des US-Außenministeriums". Since your glossary contains only "US-Außenministerium", the translation engine might not correctly add the "s" at the end or use the wrong article.

8　**Do not include multiple translations for the same source term.** For example, in English, Amazon can be both the river and the company, while in German the river is called "Amazonas". The translation may not know which German term to choose if you include two entries for "Amazon".
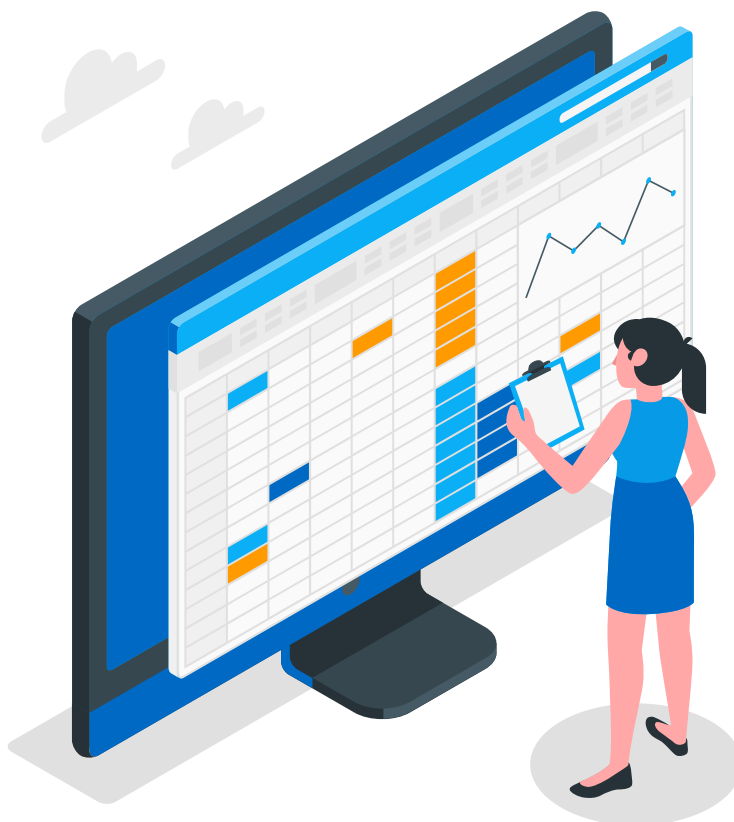
HOW TO CREATE A GLOSSARY FILE

Typically, glossaries start their lives in Excel or another spreadsheet application. Depending on the AI translation software, it may be helpful to add context to a term, including metadata such as part of speech, domain, and context. Some AI systems rely on this additional information to disambiguate terms and use them correctly in different contexts.

Your CSV file should have the source terms in the left-hand column and the target terms on the right. Here is an example of a simple CSV file in Microsoft Excel:

	A	B	C	D
1	ACME Inc.	ACME GmbH		
2	Sales department	Vertriebsabteilung		
3	ACME 2x0.5 inch Robo Widget Connectors	ACME 5x1,25 cm Robo-Dingsbumsverbinder		
4	offer	Kostenvoranschlag		
5	purchase order	Auftrag		
6				

*To save a CSV file in Microsoft Excel, go to **File > Save as**... and choose the format **CSV (comma delimited).***

⚠️ **IMPORTANT**

Some AI applications only support comma-separated CSV files. If that is the case, make sure that your CSV file uses a comma as the delimiting character, because depending on your region settings, the delimiter could also be a tab or semicolon. The easiest way to check and change the delimiter is to open your CSV file in Notepad, because Excel does not show the delimiter. For a guide on how to change the default delimiter in Microsoft Excel, see here.

```
File  Edit  Format  View  Help
ACME Inc., ACME GmbH
Sales department, Vertriebsabteilung
2x0.5 inch Robo Widget Connectors, 5x1,25 cm Robodingsbumsverbinder
offer, Kostenvoranschlag
purchase order, Auftrag
```

1 - Example of a CSV file with the correct delimiting character.

AI glossaries are best structured in a way that allows easy integration with the AI translation engine. This usually involves using standardized formats like TBX (TermBase eXchange) or CSV files. Few applications provide TBX functionality, therefore CSV has become the standard for initial glossary development. TBX is used by many translation management systems to facilitate the transfer of existing glossaries from one translation application to another.

Once integrated into the AI translation system, glossaries can automatically enforce the use of the specified terms. This helps the AI translation system to produce more consistent and accurate translations.

KEY DIFFERENCES BETWEEN AI GLOSSARIES AND HUMAN TRANSLATION GLOSSARIES

- **Usage:** AI glossaries are used by machine translation engines to enforce specific translations automatically, while human translation glossaries are used as reference tools.
- **Structure:** AI glossaries need to be in a machine-readable format with precise, unambiguous entries, while human translation glossaries can be more flexible and detailed, including contextual information and examples.
- **Flexibility:** Human translators can adapt and interpret terms based on context, whereas AI relies strictly on the glossary entries with lesser contextual understanding.
- **Updates:** Human translation glossaries often evolve through collaborative input from translators, while AI glossaries are typically developed by language engineers or terminologists to maintain system performance.

LIMITATIONS OF GLOSSARIES IN AI TRANSLATION

1

Contextual Limitations

- **Lack of Context Understanding:** AI translation systems can struggle with understanding the context in which a term is used. Glossaries provide fixed translations for terms, but without context, the AI translation system might apply the term inappropriately. For example, a word with multiple meanings might be translated incorrectly if the glossary entry does not consider the specific context in which the word appears.
- **Ambiguity and Polysemy:** Words with multiple meanings (polysemous words) can pose a challenge. Even with a glossary, an AI translation system might not correctly disambiguate between different meanings if the context isn't clear or if the glossary does not account for these nuances.

2

Integration and Maintenance Challenges

- **Technical Integration:** Integrating a glossary into an AI translation system requires a little to a lot of technical know-how and can be as easy as uploading a csv to as complex as requiring software engineering skills. Ensuring the glossary is properly formatted and compatible with the AI engine's requirements is essential, and mistakes in this process can lead to errors in translation.
- **Updates and Consistency:** Keeping the glossary up-to-date is crucial but can be labor-intensive. New terms and changes in terminology must be regularly incorporated into your glossaries, and failure to do so can result in outdated or incorrect translations. Additionally, inconsistencies between the glossary and the main translation memory can lead to errors.

3

Performance Issues

- **Processing Speed:** Large glossaries can slow down the AI translation system's processing speed. The system must check each term against the glossary, which can be time-consuming if the glossary is extensive.
- **Glossary Size and Management:** Managing very large glossaries can be cumbersome. As the glossary grows, it becomes harder to ensure that all entries are correct and consistently applied, potentially leading to translation errors.

4

Linguistic and Terminological Challenges

- **Fixed Translations:** Glossaries provide fixed translations for terms, which can limit the flexibility of the translation. This rigidity can be problematic when terms need to be adapted to fit different syntactic and grammatical contexts within sentences.
- **Terminology Overlap:** If a glossary contains terms that overlap with those already in the AI system's built-in lexicon, conflicts can arise, leading to inconsistent translations. Resolving these conflicts requires careful management and prioritization of glossary entries over built-in terms.

5 Dependency on Glossary Quality

- **Quality and Accuracy:** The effectiveness of a glossary depends heavily on the quality and accuracy of its entries. Poorly defined terms or incorrect translations can degrade the overall quality of the AI output. Ensuring high-quality glossary entries requires significant effort and expertise.
- **Coverage Limitations:** Glossaries may not cover all terms and phrases used in the source material, particularly in specialized or rapidly evolving fields. This limitation means that not all relevant terms will benefit from the glossary, potentially leading to inconsistent translations for uncovered terms.

By understanding these limitations, users can better manage their expectations and work to mitigate potential issues, ensuring more effective and accurate machine translation outcomes.

AI glossaries are used by machine translation engines to enforce specific translations automatically

WHAT ARE THE ADVANTAGES TO USING GLOSSARIES WITH MACHINE TRANSLATION?

Using glossaries with machine translation (AI) systems offers several significant advantages that enhance the quality, consistency, and efficiency of translations. Here are some key benefits:

1 Consistency and Accuracy

- **Uniform Terminology:** Glossaries ensure that specific terms are translated consistently across all documents. This is particularly important for industry-specific jargon, brand names, and technical terms, reducing the likelihood of errors and variations in translations.
- **Accurate Translations:** By providing precise translations for key terms, glossaries help maintain the accuracy of translations. This is essential in fields where precise terminology is crucial, such as legal, medical, and technical domains.

2 Efficiency and Productivity

- **Reduced Manual Effort:** Terminology errors can account for over 50% of your translation revisions. Glossaries automate the translation of frequently used terms, reducing the need for manual corrections and allowing human translators to focus on more complex and nuanced parts of the text.
- **Faster Turnaround:** With predefined translations for key terms, the AI system can process texts more quickly, leading to faster turnaround times for translation projects.

3 Enhanced Quality Control

- **Quality Assurance:** Glossaries serve as a quality control measure, ensuring that translations adhere to established terminology standards. This reduces the need for extensive post-editing and helps maintain high translation quality.
- **Error Reduction:** By standardizing term translations, glossaries help minimize the risk of mistranslations and inconsistencies, which can be costly and time-consuming to correct.

4 Adaptability and Customization

- **Customizable:** Glossaries can be tailored to specific projects, industries, or clients, allowing for customization to meet particular needs and preferences. This adaptability ensures that the AI output aligns closely with the desired terminology and style.
- **Industry-Specific Applications:** Glossaries are especially beneficial in specialized fields where precise terminology is crucial. They help ensure that the AI system uses the correct terms in the correct context, enhancing the relevance and accuracy of translations.

5 Improved Collaboration

- **Streamlined Workflow:** Glossaries facilitate better collaboration between human translators and AI systems. Translators and team members can rely on the glossary for consistent term usage, streamlining the translation process and improving overall workflow efficiency.
- **Knowledge Sharing:** Glossaries can serve as a centralized repository of terminology knowledge, making it easier for all team members to access and use the same terms consistently across projects.

6 Training and Improvement of AI Systems

- **Training Data:** Glossaries provide valuable training data for AI systems, helping them learn and apply specific terms correctly. This can improve the performance of AI engines over time, particularly in specialized domains.
- **Continuous Improvement:** As glossaries are updated with new terms and corrected translations, they contribute to the ongoing improvement of AI system accuracy and reliability.

By leveraging glossaries successfully, organizations can significantly enhance the performance of their AI machine translation systems, leading to higher quality, more consistent, and efficient translations. Successful implementation can reduce manual effort and improve translation consistency on a global scale.

Put AI-Powered Machine Translation to Work for Your Organization

Want to incorporate the power of glossaries and machine translation into your language translation workflow? **Try Pairaphrase.** It's AI-powered enterprise translation management software for organizations who value safer, smarter and faster translation.

Contact Us: pairaphrase.com/contact-us
Follow Us: @pairaphrase