A *glossary* is a custom dictionary the Cloud Translation API uses to consistently translate the customer's domain-specific terminology. This typically involves specifying how to translate a named entity.

Glossary use cases may involve:

- **Product names**: For example, 'Google Home' must translate to 'Google Home'.
- **Ambiguous words**: For example, the word 'bat' can mean a piece of sports equipment or an animal. If you know that you are translating words about sports, you might want to use a glossary to feed the Translation the sports translation of 'bat', not the translation for the animal.
- **Borrowed words**: For example, 'bouillabaisse' in French translates to 'bouillabaisse' in English. English borrowed the word 'bouillabaisse' from French in the 19th century. An English speaker lacking French cultural context might not know that bouillabaisse is a fish stew dish. Glossaries can override a translation so that 'bouillabaisse' in French translates to 'fish stew' in English.

The maximum number of glossaries for a GCP project is 1000. If you reach this limit, you can delete your existing ries (#delete-glossary) to create new ones.

The terms in a glossary can be single tokens (words) or short phrases (usually fewer than five words). The current limit for the number of separate glossaries is 1000 per project.

The main steps for using a glossary are:

1. Create a glossary file (#format-glossary)
2. Create the glossary resource (#create-glossary) with our Cloud Translation API
3. Specify which glossary to use when you request a translation (#v3)

A project can have multiple glossaries. You can get a list of the available glossaries (#list-glossaries) and can delete glossaries (#delete-glossary) that you no longer need.

Fundamentally, a glossary is a text file in which each line contains corresponding terms in multiple languages. The Cloud Translation API supports both *unidirectional* glossaries, which specify the desired translation for a single pair of source and target languages, and *equivalent term sets*, which identify the equivalent terms in multiple languages.

The total number of terms in an glossary input file can't exceed 10.4 million (10,485,760) UTF-8 bytes for all terms in all the languages combined. Any single glossary term must be less than 1024 UTF-8 bytes. Terms longer than 1024 bytes are ignored.

The Cloud Translation API accepts TSV, CSV or TMX files.

The Cloud Translation API supports tab-separated files, where each row has this format:

- `Term in source language` *tab* `Term in target language`

For example:

The tab-separated source data does not include language codes to identify the source and target languages. You identify the source and target language codes when you create the online glossary (#create-glossary).

CSV files have the same format as TSV, except a comma (",") is used as a column separator.

Translation Memory eXchange (TMX) is a standard XML format for providing source and target translations. The Cloud Translation API supports input files in a format based on TMX version 1.4. This example illustrates the required structure:

The <header> element of a well-formed TMX file must identify the source language using the `srclang` attribute, and every <tuv> element must identify the language of the contained text using the `xml:lang` attribute. You identify the source and target languages using their iso-639-1 (https://en.wikipedia.org/wiki/List_of_ISO_639-1_codes) codes.

All <tu> elements must contain a pair of <tuv> elements with the same source and target languages. If a <tu> element contains more than two <tuv> elements, the Cloud Translation API processes only the first <tuv> matching the source language and the first matching the target language and ignores the rest. If a <tu> element does not have a matching pair of <tuv> elements, the Cloud Translation API skips over the invalid <tu> element.

The Cloud Translation API strips the markup tags from around a <seg> element before processing it. If a <tuv> element contains more than one <seg> element, the Cloud Translation API concatenates their text into a single element with a space between them.

If the file contains XML tags other than those shown above, the Cloud Translation API ignores them.

If the file does not conform to proper XML and TMX format – for example, if it is missing an end tag or a <tmx> element – the Cloud Translation API aborts processing it. The Cloud Translation API also aborts processing if it skips more than 1024 invalid <tu> elements.

The Cloud Translation API only accepts CSV in the GTT (Google Translate Toolkit) format.

To define equivalent term sets, you create a multi-column CSV file in which each row lists a single glossary term in multiple languages.

The first row in the file is a header row identifying the language for each column, using its iso-639-1 (https://en.wikipedia.org/wiki/List_of_ISO_639-1_codes) or BCP-47 (https://tools.ietf.org/html/bcp47) language code. You can also include optional columns for part of speech (**pos**) and a description (**description**).

The algorithm does not currently use **pos** information, and specific **pos** values are not validated.

Each subsequent row contains equivalent glossary terms in the languages identified in the header. You can leave columns blank if the term is not available in all languages.

	A	B	C	D	E	F
1	**en**	**es**	**zh-Hans**	**ar**	***pos***	***description***
2	account	cuenta	帐户	حساب	noun	A user's account. Do not use as a verb.
3	directions	indicaciones	路线	الاتجاهات	noun	In the sense of how to get from point a to point b.
4	favorite	favorito	收藏	مفضلة	noun	Used as a noun, refers to something that is preferred.

Once you have the equivalent glossary terms identified, make the glossary available to the Cloud Translation API by creating the online glossary.

ed permissions:

oud Translation API requires that you grant **`cloudtranslate.glossaries.create`** permission to your service acc r to create a glossary. To do this, grant a role to your service account using one of the pre-defined IAM roles slate/docs/intro-to-v3#iam), such as **Cloud Translation API Editor (roles/cloudtranslate.editor)**, or you can create n role (/iam/docs/understanding-custom-roles) for your GCP project that grants the

translate.glossaries.create permission. The legacy project roles, Editor and Owner, also grant this permissio
heir wide scope, they are not recommended.

ormation on adding an account to a role, see Granting, changing, and revoking access to resources
/docs/granting-changing-revoking-access#grant_access). You can view the list of all Cloud Translation API permissi
M permissions reference (/iam/docs/permissions-reference). Cloud Translation permissions begin with
translate.".

When creating a unidirectional glossary, you must indicate the language pair (`language_pair`), by specifying the source language (`source_language_code`), and the target language (`target_language_code`).

Once you have the glossary terms identified in your equivalent term set, make the glossary available to the Cloud Translation API by creating the online glossary.

Creating a glossary is a long-running operation, so it may take a substantial amount of time to complete. You can poll the status of this operation to see if it has completed, or you can cancel the operation.

For more information, see Long-running operations (/translation/docs/advanced/long-running-operation).

In Cloud Translation - Advanced, you explicitly specify which translation model (/translation/docs/advanced/translating-text-v3#nmt-or-pbmt-model) to use for translating the text. You can also identify a glossary to use for domain-specific terminology.

The specified glossary must be associated with the current Google Cloud project.

A project can include numerous glossaries. This section describes how to retrieve a list of the available glossaries for a particular project.

The following example deletes a glossary.