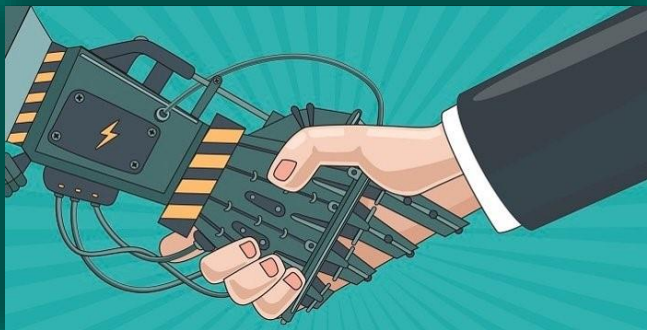


*hi!*



—  
**NMT plus a bilingual glossary:  
does this really improve terminology  
accuracy and consistency?**  
—

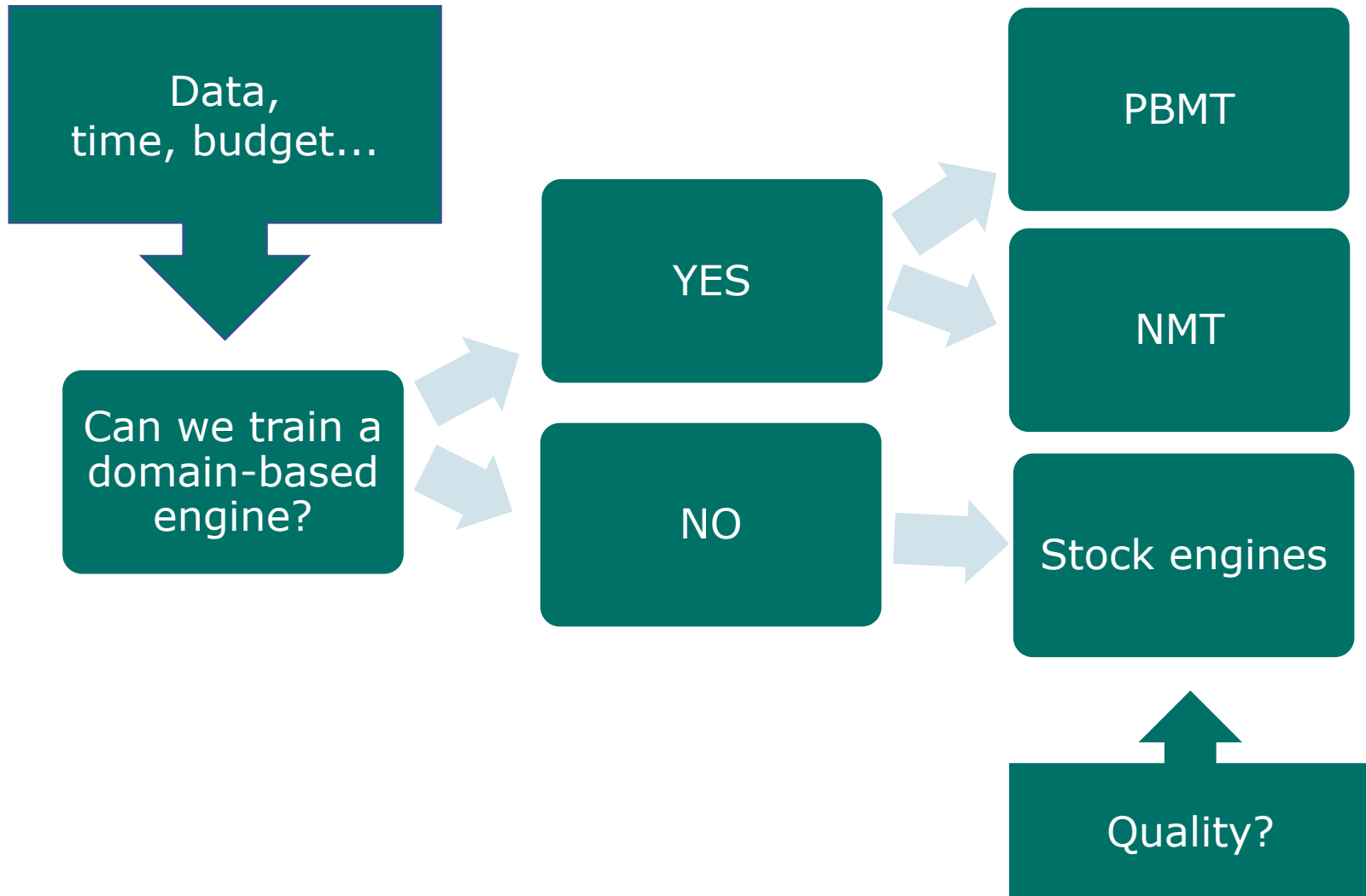


CPSL

**Lucía Guerrero**  
**Machine Translation Specialist**  
lguerrero@cpsl.com

**TC42 Asling, November 2020**

# MT at CPSL: finding the best choice



## Human assessment: A/F evaluations

«Overall, the machine translation is suitable for post-editing. However, several problems were detected as regards **terminological consistency.**»

«The MT is suitable for post-editing. The most noticeable issue is the **inaccuracy of technical terms.**»

«Most of the strings required edition, but the main reason was for **terminology issues.** Therefore, once terminology is established, I think this would reduce the number of segments that would require editing.»

«Quality of MT is good for full post-editing. Main issues are **terminology-related.** It would be a good idea to incorporate a master glossary when running MT if possible.»

# Human assessment: MT feedback report

## MT raw output feedback

Ref.	Source	Raw MT output	Post-edited text	Error Category (drop-down menu)	Severity (drop-down menu)	Comments
	baloon	globo	balón	terminology	3	
	adverse events	eventos adversos	acontecimientos adversos	terminology	2	
	IAB	IAB	BIA	other	2	acronym
	IABP	IABP	CBIA	other	2	acronym
	Every 2 hours automatic recalibration	Cada 2 horas de recalibración automática	Recalibración automática cada 2 horas	language	3	
	bench resting	pruebas de banco	pruebas comparativas	style	2	
	Data on file	Datos en archivo	Datos de archivo	language	3	

# NMT stock engines plus glossary



## How do they work?

### Creating and using glossaries (Advanced)

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 Cloud Translation API the sports translation of "bat", not the translation for the animal.

**Borrowed words:** For example, "bouillabaisse" in French translates to "bouillabaisse" in English (...) Glossaries can override a translation so that "bouillabaisse" in French translates to "fish stew" in English.

Source: <https://cloud.google.com/translate/docs/advanced/glossary>

## How do they work?



### Best Practices

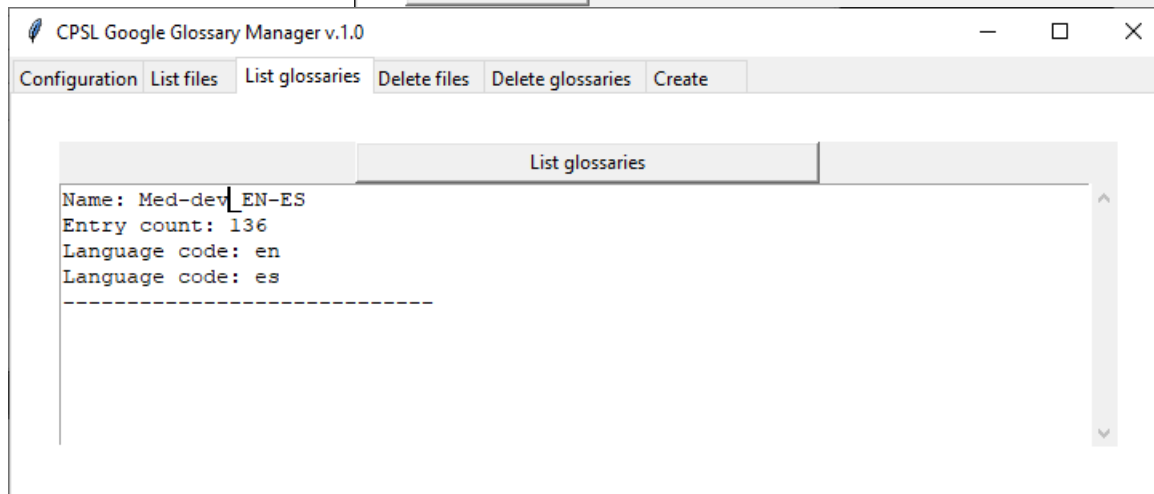
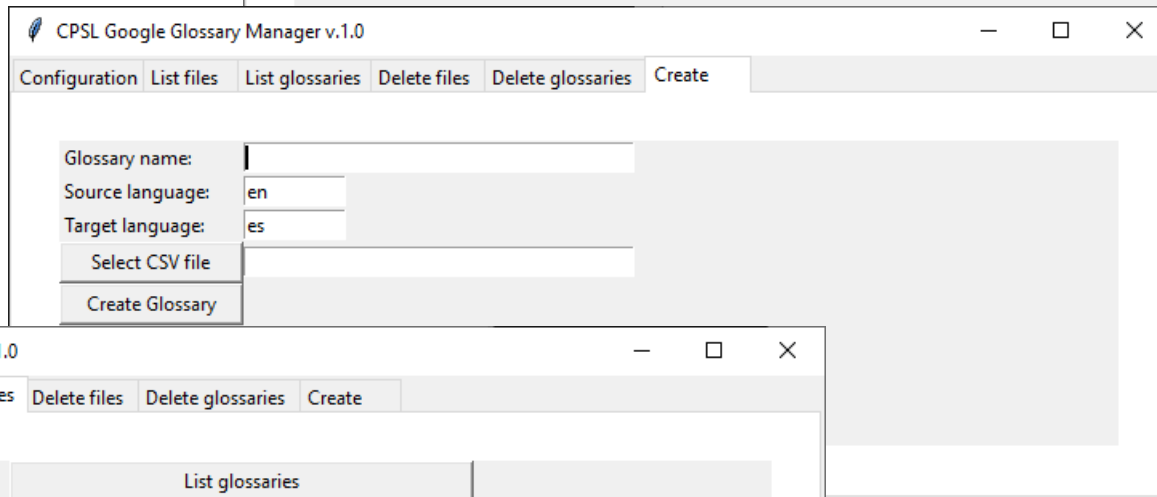
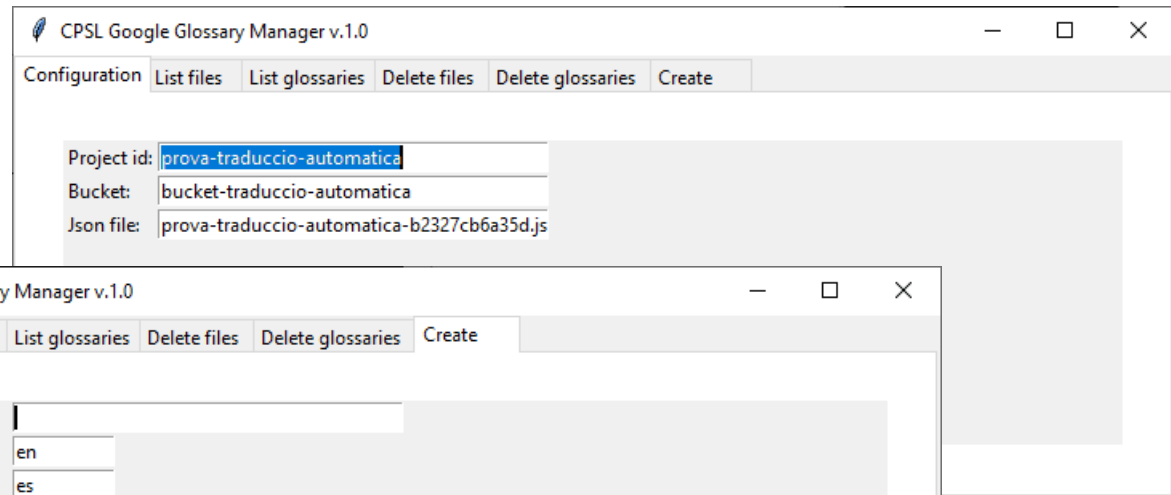
- **Keep your custom terminology minimal.** Only include words which you want to control and which are completely unambiguous. Only use words that you know you will never want to use an alternate meaning of, and you want it to only ever be translated in a single way. Ideally, limit the list to proper names, like brand names and product names.
- **Custom terminologies are case-sensitive.** If you need both capitalized and non-capitalized versions of a word to be included, you must include an entry for each version.
- **Do not include different translations for the same source phrase** (for example, entry #1-EN: Amazon, FR: Amazon; entry #2-EN: Amazon FR: Amazone).
- Some languages do not change the shape of a word based on sentence context. With these languages, applying a custom terminology is most likely to improve overall translation quality. However, **some languages do have extensive word shape changes.** We do not recommend applying the feature to those languages, but we do not restrict you from doing so.

Source: <https://docs.aws.amazon.com/translate/latest/dg/ct-best-practices.html>

# Integration into CAT tools



Universitat  
Oberta  
de Catalunya







**Medical devices**



**English into Spanish (Spain)**

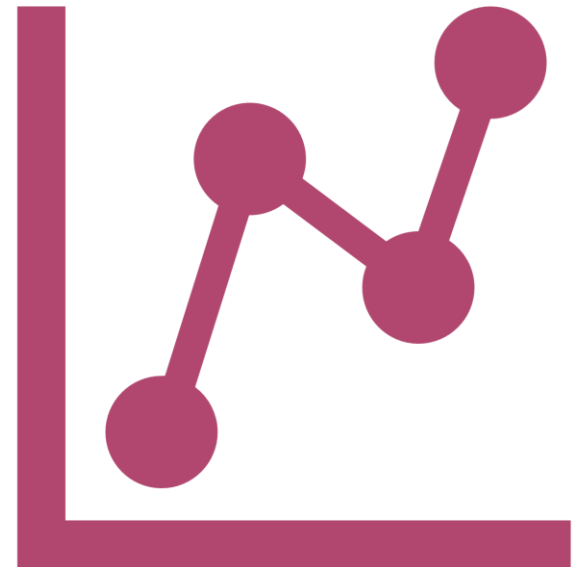


**High quality expected  
(for publishing)**



**Raw MT + full PE + review**

- 4 translation requests
  - Edit distance (Levenshtein algorithm from nltk.metrics)
  - Comparison reports
  - Human analysis
  
- 2 MT systems:
  - Google Translate Basic
  - Google Translate Advanced with a glossary



## — Our findings



### **Glossary term**

dialysis access = diálisis

### **Source**

Premium performance for **dialysis access**

<b>Google Translate Basic</b>	<b>Google Translate Advanced</b>
Rendimiento superior para * <b>acceso de diálisis</b>	Rendimiento superior para <b>diálisis</b>

## Glossary term

IAB = BIA (balón intraórtico)

## Source

**IAB** Minor Vascular Complications

Google Translate Basic	Google Translate Advanced	Final version
* <b>IAB</b> complicaciones vasculares menores	<b>BIA (balón intraórtico)</b> complicaciones vasculares menores	Complicaciones vasculares menores con <b>BIA (balón intraórtico)</b>

—  
**Our findings:  
number**

## **Glossary term**

IAB = BIA (balón intraórtico)

## **Source**

Higher efficacy fiber-optic **IABS**

<b>Google Translate Basic</b>	<b>Google Translate Advanced</b>	<b>Final version</b>
* <b>IABS</b> de fibra óptica de mayor eficacia	* <b>IABS</b> de fibra óptica de mayor eficacia	<b>BIA</b> de fibra óptica de alta eficacia

—  
**Our findings:  
number**

**Glossary term**

graft = prótesis

grafts = prótesis

**Source**

Survival (%) of patients with central venous catheters versus arteriovenous fistulas/**grafts**

**Google Translate Basic**

Supervivencia (%) de pacientes con catéteres venosos centrales versus fístulas / **\*injertos** arteriovenosos

**Google Translate Advanced**

Supervivencia (%) de pacientes con catéteres venosos centrales versus fístulas / **prótesis** arteriovenosas

## Glossary term

early stick graft = prótesis de canulación temprana

early stick grafts = prótesis de canulación temprana

## Source

“**Early stick grafts**” may offer elderly patients the option to avoid CVCs with their inherent “high risk of infection”

### Google Translate Basic

Los **\*“injertos de palo tempranos”** pueden ofrecer a los pacientes de edad avanzada la opción de evitar los CVC con su inherente “alto riesgo de infección”

### Google Translate Advanced

**\*Los “prótesis de canulación temprana”** pueden ofrecer a los pacientes de edad avanzada la opción de evitar los CVC con su inherente “alto riesgo de infección”

—  
**Our findings:  
ambiguity/polysemy**

**Glossary term**

lift = elevador

**Source**

The tissue-conforming suction cup uses a gentle vacuum to securely **lift** and hold the heart.

**Google Translate Basic**

La ventosa que se adapta a los tejidos utiliza una aspiradora suave para **elevantar** y sujetar el corazón de forma segura.

**Google Translate Advanced**

La ventosa que se adapta a los tejidos utiliza una aspiradora suave para **\*elevador** y sujetar el corazón de forma segura.

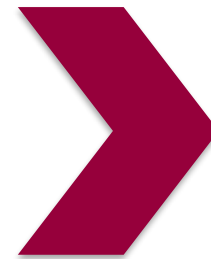


# Edit distance

Google Translate Basic	<b>EdDist%</b>			
	31.99	<b>Projects &gt; 25% EdDist:</b>	<b>54%</b>	<b>7 out of 13</b>
	30.32	<b>Projects &gt; 30% EdDist:</b>	<b>23%</b>	<b>3 out of 13</b>
	28.59	<b>Average EdDist:</b>	<b>25%</b>	
	26.58			
	25.47			
	20.49			
	13.24			
	28.59			
	20.49			
	20.61			
	19.81			
	23.53			
	33.94			
GT Advanced	18.65	<b>Projects &gt; 25% EdDist:</b>	<b>50%</b>	<b>2 out of 4</b>
	17.3	<b>Projects &gt; 30% EdDist:</b>	<b>25%</b>	<b>1 out of 4</b>
	26.87	<b>Average EdDist:</b>	<b>24%</b>	
	32.37			

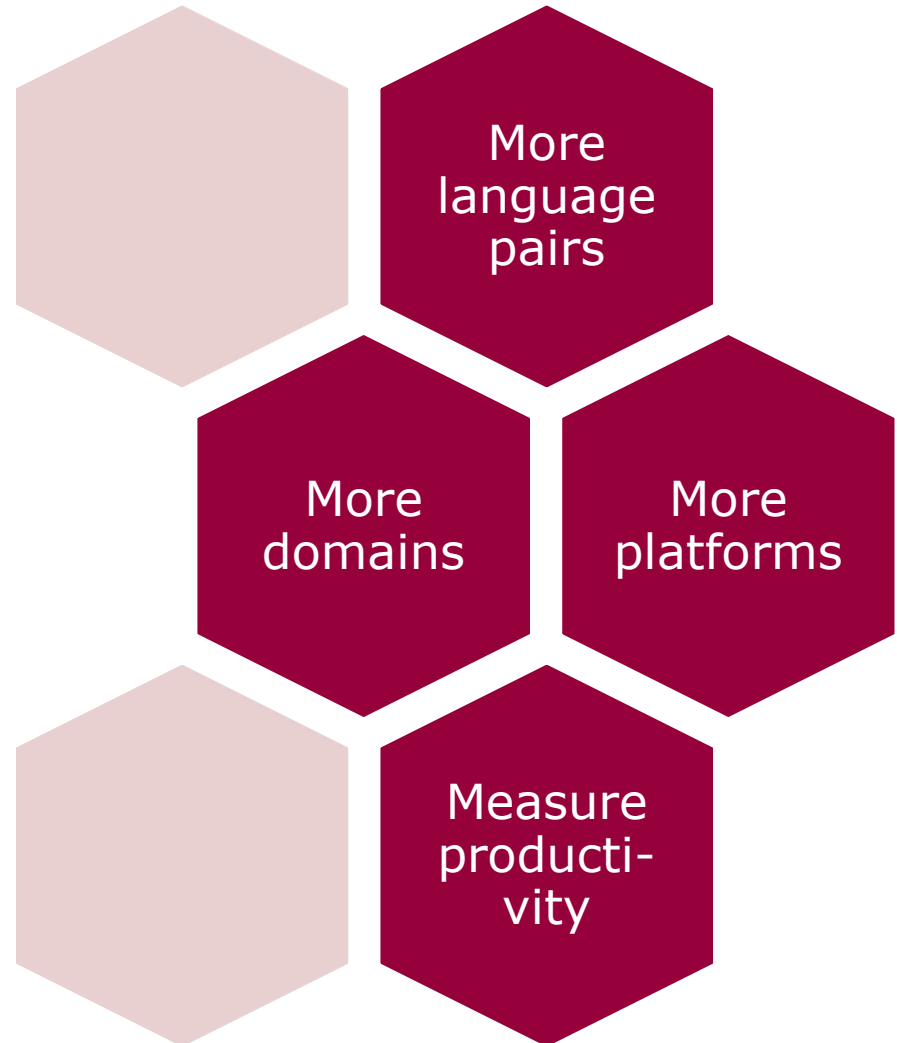
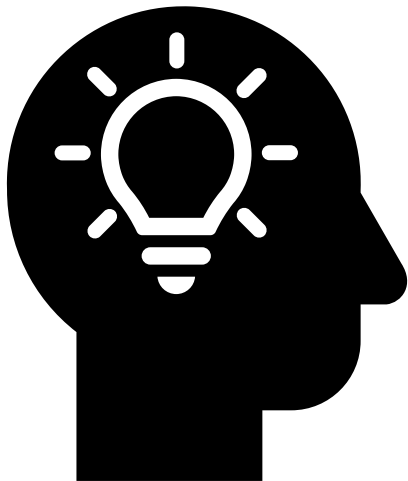
## Conclusions

- Requires pre-processing of existing termbases or creating *ad hoc* glossaries
- Reduces inconsistencies
- Works best when translating into less rich morphologically languages
- Works best with not translatables such as brand/product names
- Can create new errors



**Interesting...  
but not panacea  
Use with caution!**

# Ideas for further study



*Questions?*

**Thank you!**



## **CPSL Barcelona**

Tel +34 93 320 13 00  
info-spain@cpsl.com

## **CPSL Madrid**

Tel +34 93 320 13 00  
info-spain@cpsl.com

## **CPSL Germany**

Tel +49 (0)71 41 - 97 00 55  
info-germany@cpsl.com

## **CPSL UK**

Tel (+44) 207 993 4550  
info-uk@cpsl.com

## **CPSL USA**

Tel (+1) 214 432 6114  
info-usa@cpsl.com

**cpsl.com**

**CPSL**  
Language Services

LANGUAGE  
**AS**  **ING**  
TECHNOLOGY

**Lucía Guerrero**  
**Machine Translation Specialist**  
lguerrero@cpsl.com