



Leveraging Translations without a Workable Source (aka "orphan" Documents)

Jean-Marie Guyot

Cat.NIP Solutions | *Nifty Improved Process for CAT*



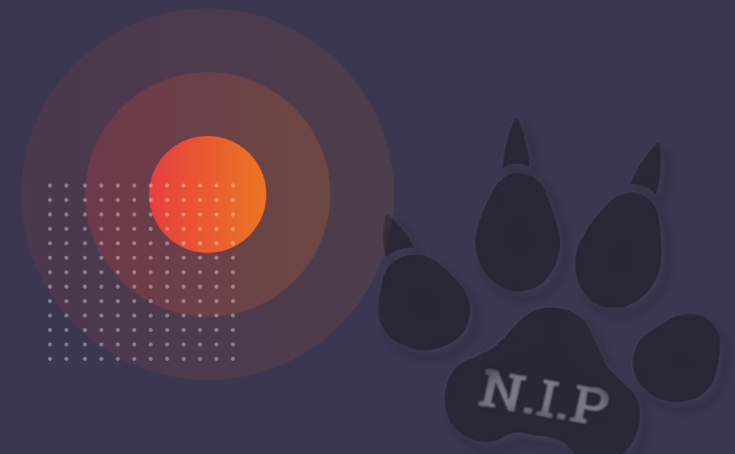
Re-using is more efficient than translating from scratch

The obvious benefits of re-using existing translations include:

- **Cost and Speed:**
smaller volumes to translate = faster and cheaper

But its most important benefit may be:

- **Consistency/Quality:**
don't change things that have already been approved
preserve the familiar style and terminology



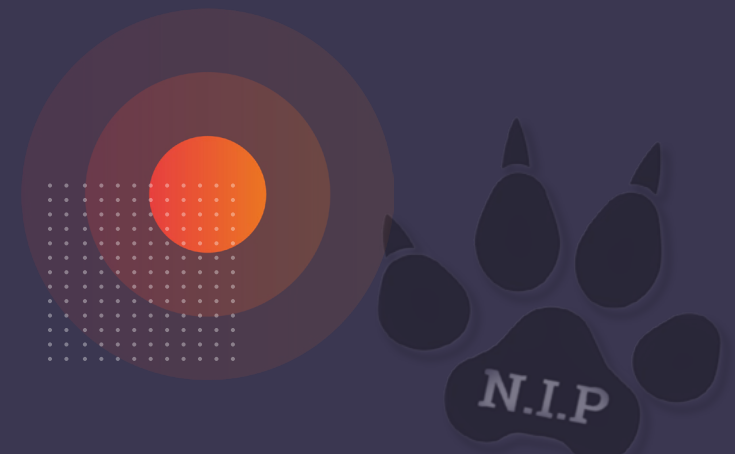
Why do we use TMs?

We use Translation Memories to:

- re-use translations (unchanged source, aka 101% or 102%)
- recycle translations (identical or similar source, 100% and fuzzies)

It's a very simple concept used in all CAT systems.

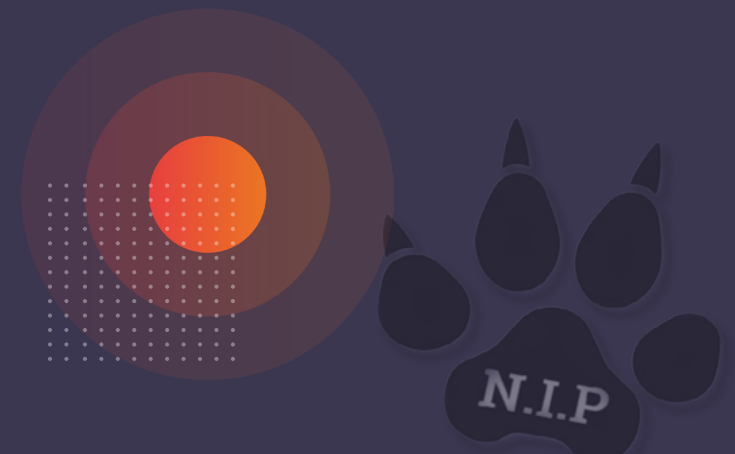
The main benefit of a TM is to provide translations that have **already been used/approved** in a similar environment.



If we don't have TMs

Without TMs, we can:

- Translate from scratch
with or without the help of advanced tools...
(this topic will be no doubt be discussed in many other sessions at mQfest)
- Try to leverage existing references
align older translations
extract **terminology** and make **glossaries**
use target-concordance in LiveDocs
read and get inspired...



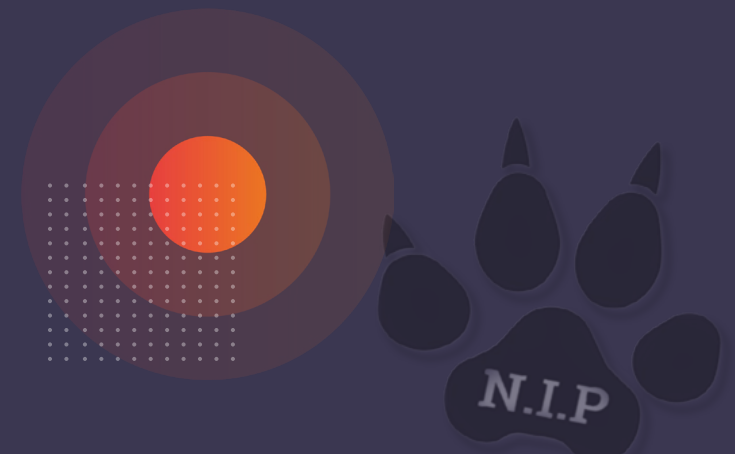
No reference at all

Translating from scratch is perfectly fine when there are no existing material (either previous translations of existing documents published by the local office).

There are many tools available to boost the translator's job.

No reference = no risk of being inconsistent.

Our problem is when we have existing material, but with no simple way of re-using it.



Pros & Cons of aligning existing translations

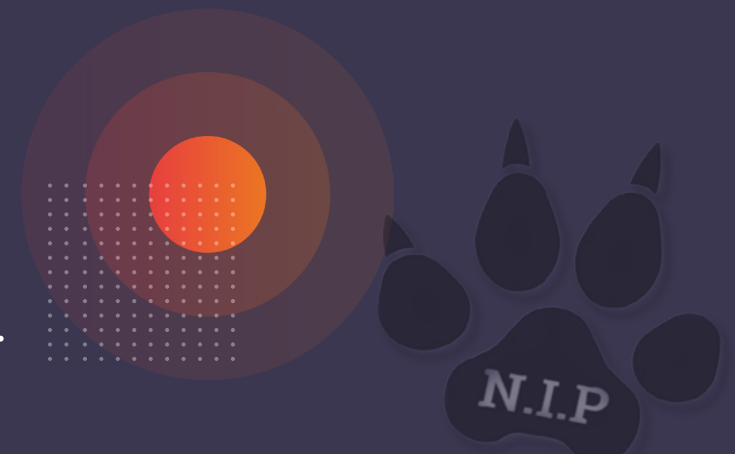
The Cost & Speed benefits of alignment can be challenged:

- It is time-consuming
- It can be expensive
- It may deliver unreliable results
- You don't know if the final results will be needed
- ... It doesn't always work

Note: the dynamic approach of memoQ LiveDocs can solve some of these issues. But the alignment process itself remains very limited 😞.

But the Consistency benefit remains strong, since all existing translations were previously approved.

So, alignment is great ... but only when simple and straightforward.



4 questions to leverage your existing translations



Do you have TMs or bilingual docs?

YES: Use them to translate → [BEST]

NO: ...?



Do you have existing translations / local docs?

NO: Translate from scratch → [ALL NEW]

YES: ... you have to remain consistent with existing! ...



Do you have workable source documents?

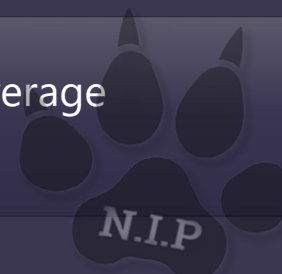
YES: Cross your fingers and start aligning → [GREAT BUT TRICKY]

NO: Read the existing docs before translating → [WORST]



Do you know **ShadowMatching**?

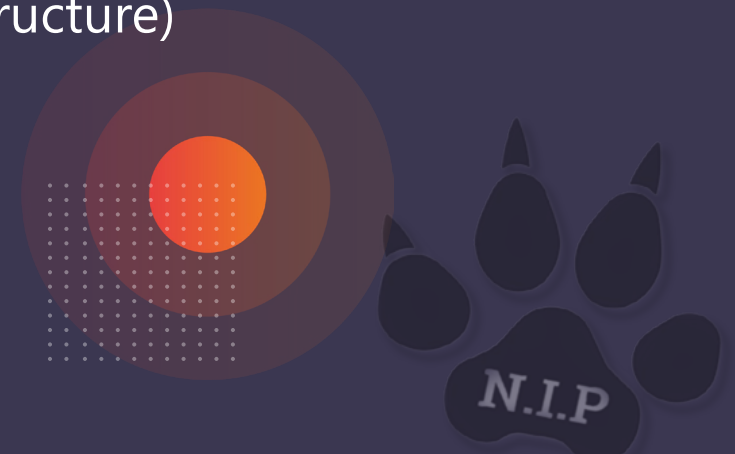
ShadowMatching is a simple process to safely leverage **all existing** target references → [QUICK & SAFE]



What do we really **want**?

If we could dream of a solution, we would want it:

- Fast
- Affordable
- Suggests only **approved translations** (no MT guesses)
- Can leverages **all** existing references (translations and also local writing)
- No chance of misalignment (even with complex document structure)



Can Machine Translation help?

We are not looking for *translations* from an MT system.

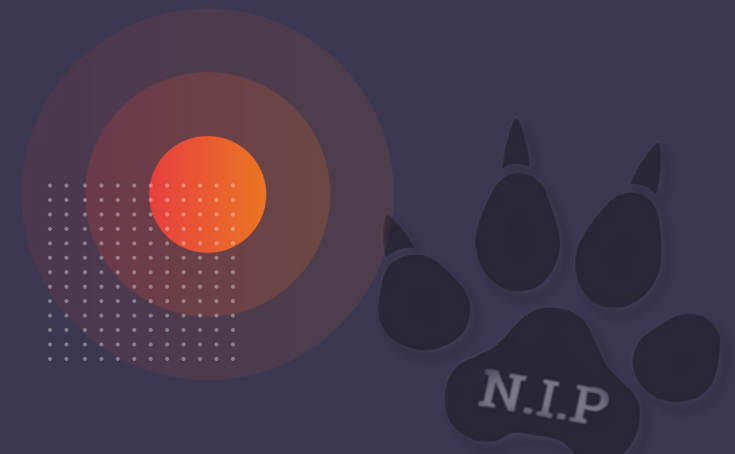
But we want bilingual segments that will **suggest approved sentences** coming from our reference pool of local writing and orphan translations.



Let's **flip the table** and use machine translation...

... to generate **the most likely source(s)** for each sentence

Let's start **ShadowMatching**
ShadowMatching

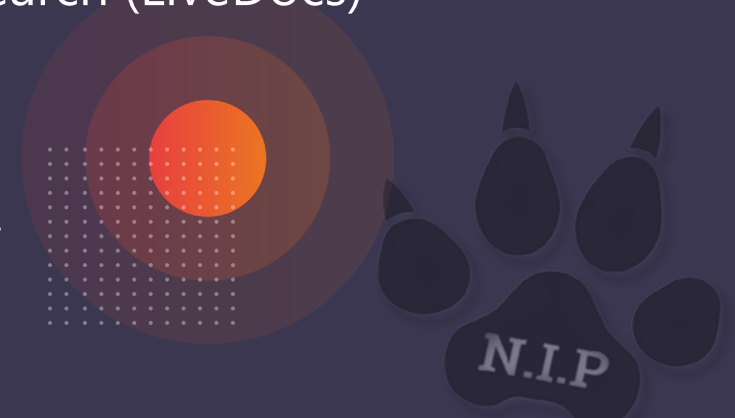


The ShadowMatching principles

- Use MT to **guess** what the **source** of an existing translation *was or could have been* for each locally written material
- Make all these guesses a translation resource (LiveDocs or TM)
- No need to verify the alignment
- If we have a match, it provides only **real translations** or local content
- If our MT-guessed source is wrong... do nothing
- The entire pool of reference text remains available for context search (LiveDocs)

Important:

Correct segmentation of the imported reference file is paramount.
This is the key factor of all CAT process.



ShadowMatching Process



Segment the target references

- Load all reference documents into memoQ to create a pool of **translated** segments

MT Translate these segments into your missing "source" language

- You can use one or several MT engines to increase chances of high matches

Flip the translation units

- Swap the source and target data

Load the flipped bilingual doc into a LiveDocs or a TM

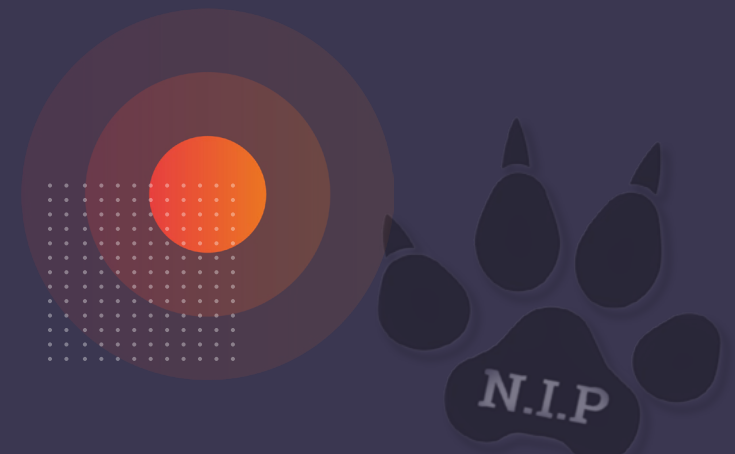
- If the MT reverse-translation is close to the original source, you will get a match with the correct translation.
- Make sure that they are clearly identified as ShadowMatch

A few points about ShadowMatching

- You can use multiple MT engine to increase your chances of matching
 - Terminology-based MT is recommended as it will deliver more relevant matches
- A wrong guess is not a problem; the translation will just not be suggested
- Native, non-translated documents can also be used
- The only limit for reference material is the segmentation status
 - Bad segmentation will deliver bad matches

Recommendations

- Load ShadowMatched references into a dedicated LiveDocs to differentiate it from real TMs
- Mark (pollute) the shadow source to quickly identify it





QUICK DEMO
Leveraging orphan
target text with
ShadowMatcher
(or later if time is short)



thankQ!

I have one double question for you:

- Have you been in this situation?
--- and what did you do?

Your Questions:

- That's your turn...

jm.guyot@catnipsolutions.com

Cat.NIP Solutions | *Nifty Improved Process for CAT*