Naoko Uchida

Translation Tools: Friend or Foe?
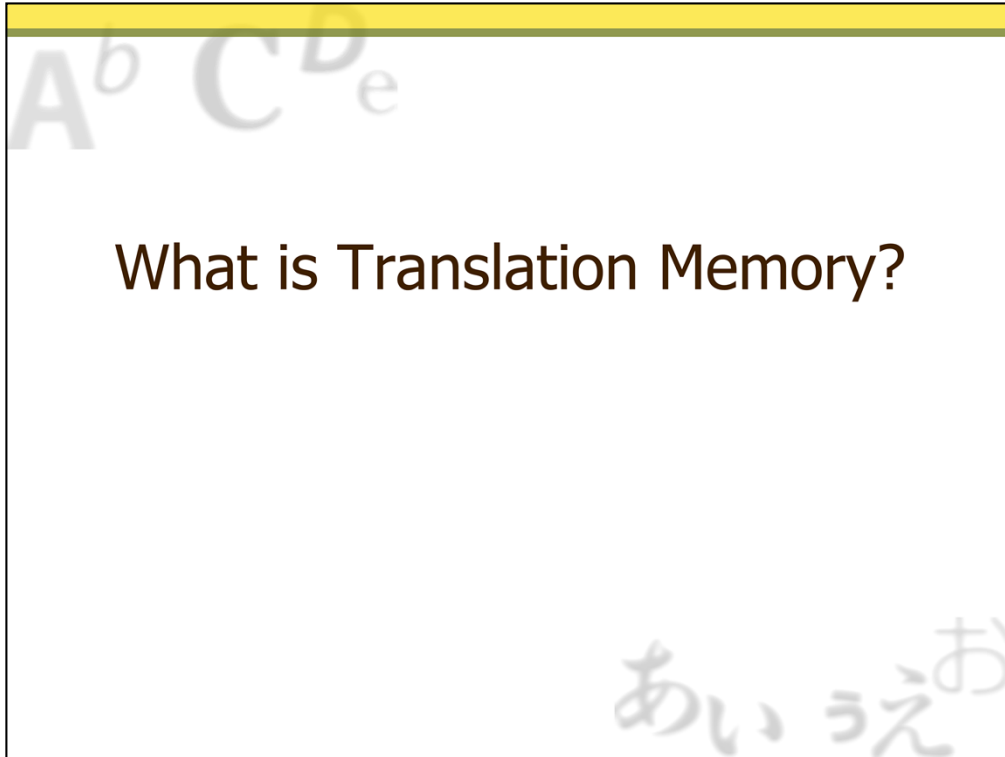
IJET-22 Seattle

Content

- What is Translation Memory?
- Changes in TM Tools and Surrounding Environment
- TM Tool as a Differentiation Factor for Translators
- Conclusion

What is Translation Memory?

To start, I would like to quickly go over What a Translation Memory is, for those who have not worked with it.

What is Translation Memory?

- Database that stores "segments" that have been previously translated.

A translation memory, or TM, is a database that stores "segments" that have been previously translated.

"Segments" can be sentences or sentence-like units, such as headings, titles or elements in a list.

What is Translation Memory?

- Database that stores "segments" that have been previously translated.

- Typically used in conjunction with a dedicated Computer Assisted Translation (CAT) tool.

Translation memories are typically used in conjunction with a dedicated computer assisted translation tool.
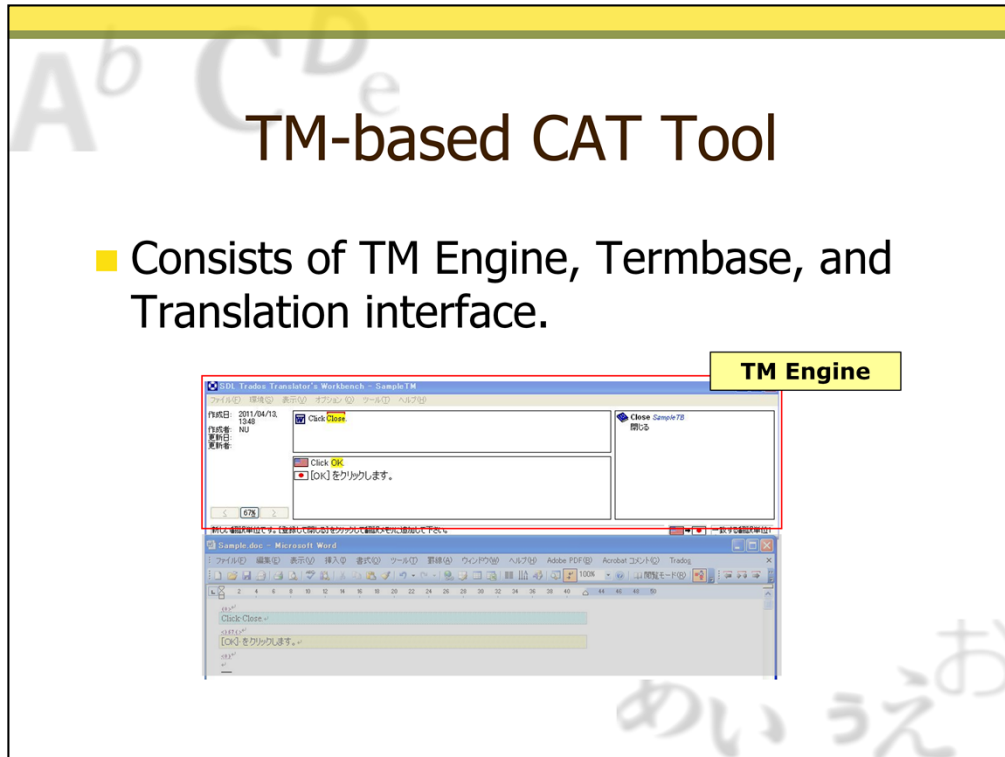
TM-based CAT Tool

- Consists of TM Engine, Termbase, and Translation interface.

The TM-based translation tool typically consists of 3 components.

TM Engine, Termbase, and Translation interface. I would like to briefly go over each components.
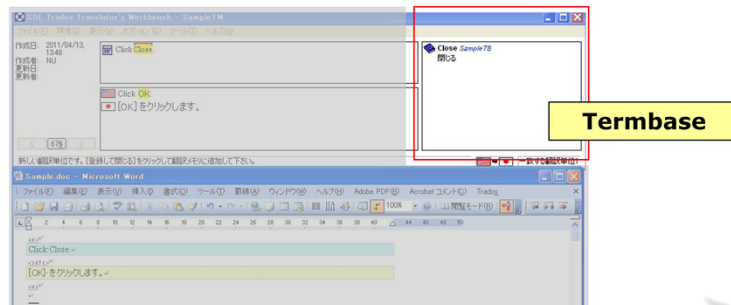
TM-based CAT Tool

- Consists of TM Engine, Termbase, and Translation interface.

I use a screen shot from SDL Trados 2007 as an example. TM Engine provide access to Translation Memories.

TM-based CAT Tool

- Consists of TM Engine, Termbase, and Translation interface.

As for Termbase, what is shown here is not the actual termbase. But it is a viewer window that displays relevant terminologies that are stored in the termbase. Termbase is maintained with a separate program which is generally called terminology management systems.

TM-based CAT Tool

■ Consists of TM Engine, Termbase, and Translation interface.
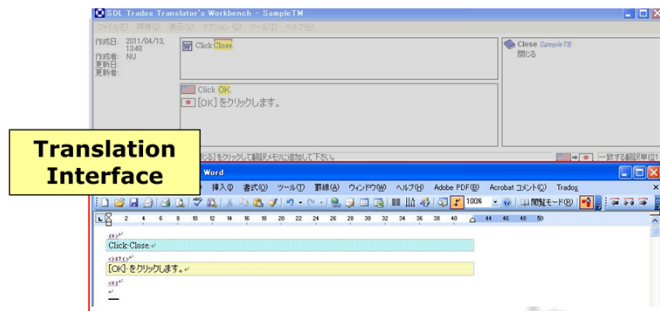
Translation interface is the area where the translator types in her translation. In this example, it is Microsoft Word. This TM Engine, SDL Trados Translator's Workbench, can work in conjunction with MS Word. Other CAT tools may have a translation interface integrated in a single interface window with TM Engine and Termbase.

TM-based CAT Tool

- Consists of TM Engine, Termbase, and Translation interface.
- **Breaks the source text into segments.**

Now I move on to how TM-based CAT tools assist translation. First the program breaks the source text, that is the text to be translated, into segments.

TM-based CAT Tool

- Consists of TM Engine, Termbase, and Translation interface.
- Breaks the source text into segments.
- Matches between the segments and the previously translated source-target pairs.

Then, it looks for matches between those segments and the source half of previously translated source-target pairs stored in a translation memory.

TM-based CAT Tool

- Consists of TM Engine, Termbase, and Translation interface.
- Breaks the source text into segments.
- Matches between the segments and the previously translated source-target pairs.
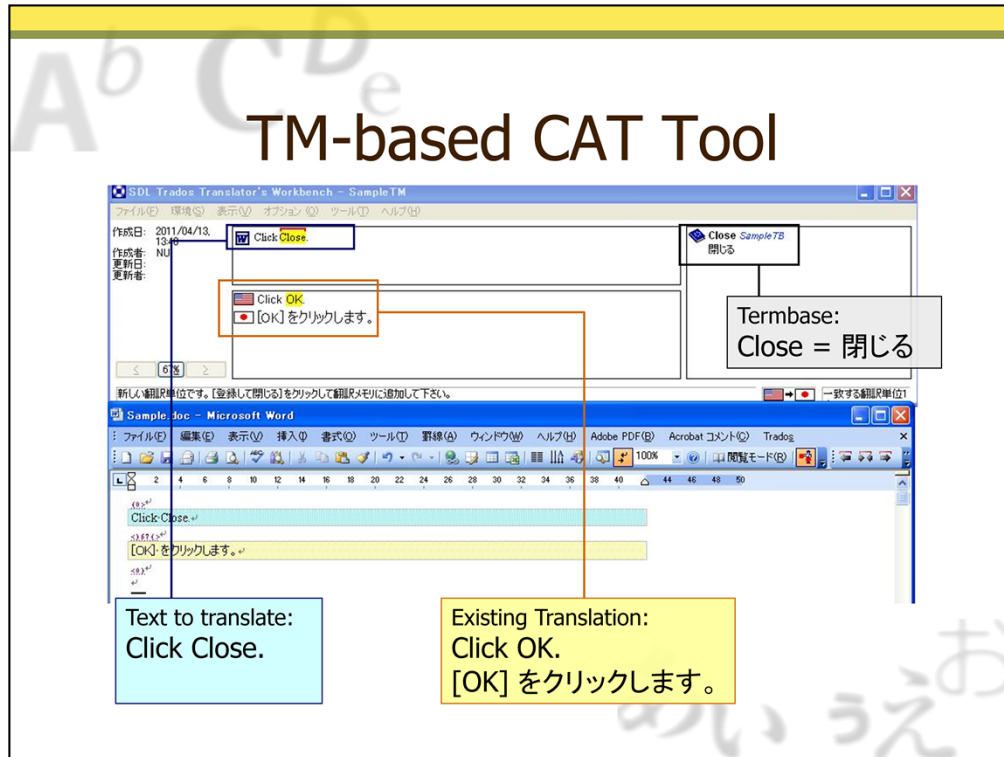- **Presents matching pairs as translation candidates.**

Then, it presents such matching pairs as translation candidates. The translator can accept a candidate, replace it with a fresh translation, or modify it to match the given source segment. The new or modified translation goes into TM as a newly translated source-target pair.

In this example, the text to be translated is "Click Close." <click1>

The TM Engine finds a close match, "Click OK." <click2> that has already been translated and stored in the TM.

The word "Close" and "OK" is highlighted to indicate where the current source segment is different from the source half of the matching pair. <click3>

In this case, the word "Close" has a entry in the Termbase, so the relevant entry is displayed in the Termbase window. <click4>

By simply replacing the "OK" in the translation candidate with the target termbase entry for the word "Close", translation is done.

In this way, the translator can translate a given text more quickly and efficiently than she would without the CAT tool.

The tool also helps to maintain consistency in style and terminology use within or across documents.

Changes in TM Tools and Surrounding Environment
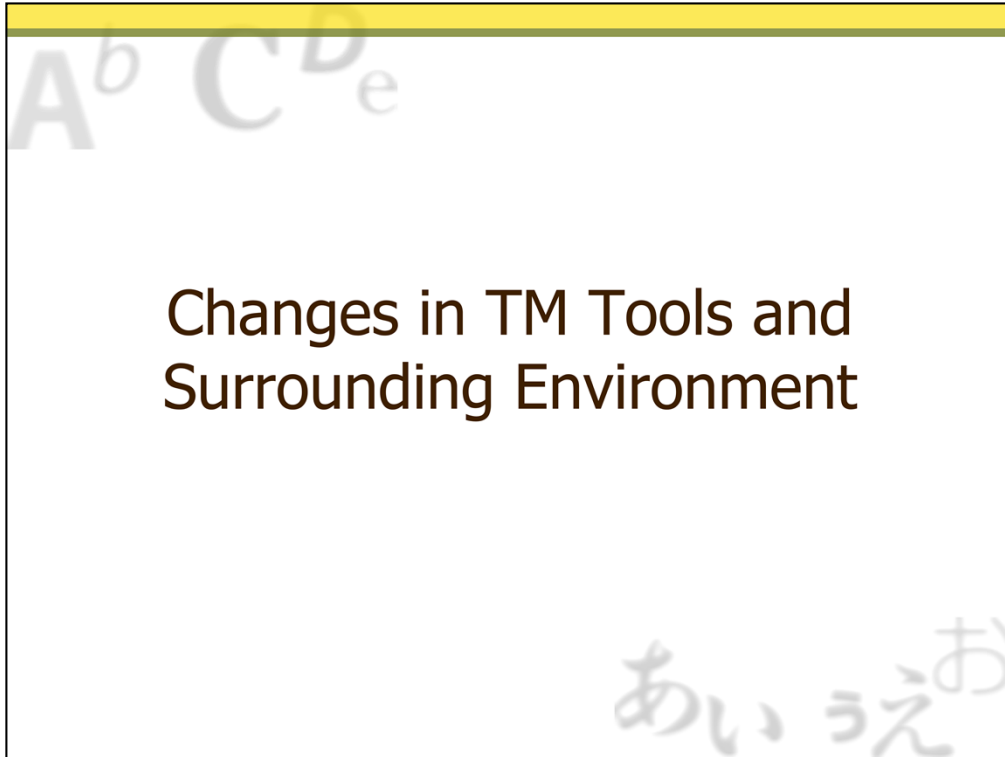
I would like to share my view of Changes in TM Tools and Surrounding Environment.

1997-2000: Early Period

- CAT tool was often not a requirement.

Let's start with 1997-2000: Early Period.

In those days, only handful of companies and translation agencies used them.

Because few people had it, it was often not considered as a requirement.

1997-2000: Early Period

- CAT tool was often not a requirement.
- My own productivity tool.
  - More output with less time. More $.

Although I had to use my own time&effort to create TM from scratch,

→TM allowed me to output more in less time. This was enough to offset the startup time.

What's great was that I get to charge full word counts for everything I translate, including those words translated the way I just showed you in previous slides.

1997-2000: Early Period

- CAT tool was often not a requirement.
- My own productivity tool.
  - More output with less time. More $.
- **TM is maintained at translator's side.**
  - Customized for precision and efficiency.
  - Utilized attribute feature:

ST → T1 <MainText>, T2 <IndexText>, T3 <UI>
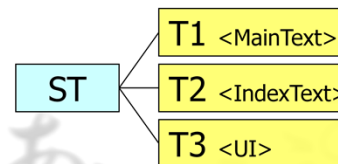
I created my own TM. I also did maintenances on my own. And this served to my benefit as well.

Having a full control over the TM, I was able to apply a range of customizations to it for better precision and efficiency.

In default setting, Trados only takes one version of translation per source sentence, because Translation memory saves translation as 1-to-1 pair with the original.

I will not discuss this in detail, but I did a presentation on this subject at ATA conference in the past.

If any of you are interested, the presentation material is still available to view from my website.

1997-2000: Early Period

- CAT tool was often not a requirement.
- My own productivity tool.
  - More output with less time. More $.
- TM is maintained at translator's side.
  - Customized for precision and efficiency.
  - Utilized attribute feature:
- **File based TM**

File based TM means a translation memory that is made of files saved on your own hard drive, as opposed to server base TM (explain later).

What is important is that you have a full access and control over the whole TM.

2000-2007: Trados dominant

- Trados was required in most projects.

The next time period is 2000 to 2007. I call this period "Trados dominant."

As far as I can remember, everyone was using Trados. Even the trial translation required it.

There were some clients who did not ask to use any CAT tools, but most of agency clients required Trados to be used for their assignments.

It was not just any CAT tool. They specifically asked for Trados.

2000-2007: Trados dominant

- Trados was required in most projects.
- Efficiency tool for LSP.

And Trados has now become an efficiency tool for the agencies, who by then were often call themselves as Language Service Providers, or LSP to be short.

All the time-saving and money-making benefit has since been serving the interest of LSP, with the birth of...

2000-2007: Trados dominant

- Trados was required in most projects.
- Efficiency tool for LSP.
  - <u>Fuzzy Rate</u>: Discount for the <u>fuzzy matches</u>

X% discount	99 % match
Y% discount	98-85% match
Z% discount	84-75% match

The Fuzzy Rate.

Fuzzy rate refers to discounted rates that is applied to fuzzy matching segments.

Fuzzy matching is a technique used in computer-assisted translation. If the current source text is similar to previously translated text, it is categorized as a "fuzzy" match. Trados and many other CAT tools assign percentages to these kinds of matches, in which case a fuzzy match is greater than 0% and less than 100%. <Click1>

The LSPs who required their subcontractors to use Trados typically paid only discounted rate for the word counts with higher fuzzy match scores. <click2>

Many LSPs went even further, and apply those fuzzy matching scores to their estimated completion time.

So the translators are expected to complete the translation in shorter turn around than the traditional time frame which was based on a full word count.

In this way, translators who uses TM-based CAT tool as required by their clients are now expected to translate cheaper and faster.

2000-2007: Trados dominant

- Trados was required in most projects.
- Efficiency tool for LSP.
  - Fuzzy Rate: Discount for the <u>fuzzy matches</u>
- **TM was maintained at the client side.**

TM was maintained at the client side. Translators worked with translation memories as provided by their client.

When the translation work was complete, the translators were often asked to submit the resulting translation memory as a part of their delivery.

The translation memories were clearly considered as a part of translation assets that belong to the client.

2000-2007: Trados dominant

- Trados was required in most projects.
- Efficiency tool for LSP.
  - Fuzzy Rate: Discount for the <u>fuzzy matches</u>
- TM was maintained at the client side.
  - TM customization is restricted.
  - Less flexibility for TR → Tool translationese

This meant translator's were restricted in their ability to modify or customize existing settings in translation memories. Because such modification or customization may interfere with the TM maintenance process at the client.

Changing any TM settings may also affect the its matching statistics, which would have direct impact on the translation rate structure. For these reasons, translators were generally instructed not to touch any TM settings.

As a result, translators have lost their flexibility to modify their tools to better assist their ability to translate more naturally and precisely.

From this point on, it became more about fitting our own translation into the restriction of TM tools. And I believe this have contributed to reinforce "Tool translationese".

"Translationese" is not necessarily incorrect but awkward language produced as a result of translation.

"Tool translationese" is just a word I made up using the same analogy.

It refers to not necessarily incorrect but awkward translation produced as a result of translating with CAT tools.

I will discuss this in more detail later in this presentation.

2000-2007: Trados dominant

- Trados was required in most projects.
- Efficiency tool for LSP.
    - Fuzzy Rate: Discount for the <u>fuzzy matches</u>
- TM was maintained at the client side.
    - TM customization is restricted.
    - Less flexibility for TR → Tool translationese
- **File based TM.**

Translation memories during this time period were still mostly file based. And...

2000-2007: Trados dominant

- Trados was required in most projects.
- Efficiency tool for LSP.
  - Fuzzy Rate: Discount for the <u>fuzzy matches</u>
- TM was maintained at the client side.
  - TM customization is restricted.
  - Less flexibility for TR → Tool translationese
- File based TM.
- **Translator work with whole TM.**

Translator worked with whole TM. Translators were still provided with translation memories that included entire database of existing translation for the given project.

And you have this TM as your own copy on your own hard drive.

This allowed some flexibilities for translators to use certain utilities in conjunction with the TM to improve efficiencies.

2008-Present

- Most LSP require use of a CAT tool.

Let's move on to the present time. Most LSPs that I work with still require to use a CAT tool for their projects, but it is no longer necessary to be Trados.

CAT tools have became much more diverse now. Many LSPs have even developed their own tools to be used for their assignments.

As of now, I have 6 different CAT tools installed in my PC and I use 5 of them regularly.

2008-Present

- Most LSP require use of a CAT tool.
- A component in multilingual content management workflow.

And this diversification of CAT tool has a lot to do with the fact that it is now integrated as one component of multilingual content management workflow.

Many LSPs and any technology companies with needs to maintain substantial amount of translated contents are typically managing their translation projects through a globalization management system.

The globalization management system maintains centralized Translation memory, terminology database, workflow engine, and other related tools.

It is like the translation projects are moving along an assembly line, and translators are given access to different sections of the line depending on their role, such as translators or reviewers.

2008-Present

- Most LSP require use of a CAT tool.
- A component in multilingual content management workflow.
- **TM is maintained at the client side.**

Again, TM is maintained at the client side. And now this is reinforced much more systematically.

Translators' task with translation memories are pretty much limited to entering their translation to the translation memory.

The setting and properties of translation memories are often pre-set by the client and stay invisible to subcontracted translators while they work on those TMs.

So, translators just receive a translation memory from the client, use the TM as instructed by the client, and then submit the resulting TM along with finished translation.

Our work ends there.

2008-Present

- Most LSP require use of a CAT tool.
- A component in multilingual content management workflow.
- TM is maintained at the client side.
- **Server based TM.**

Translation memories are now server based. According to Wikipedia, the system is called Centralized translation memory systems.

Centralized translation memory systems store TM on a central server. They typically provide desktop TM engine. Translators can install it in their own PC to work offline.

2008-Present

- Most LSP require use of a CAT tool.
- A component in multilingual content management workflow.
- TM is maintained at the client side.
- Server based TM.
- **Translator work with a subset of the TM.**
  - Prebuilt "Translation Kits"

Translators are now only get to work with a subset of the project's TM.

The centralized translation memory system export prebuilt "translation kits" for translators to work with.

This "translation kit" contains content to be translated which is pre-segmented on the central server. It also contain a subset of the TM containing only applicable TM matches.

This further limits flexibility for translators.
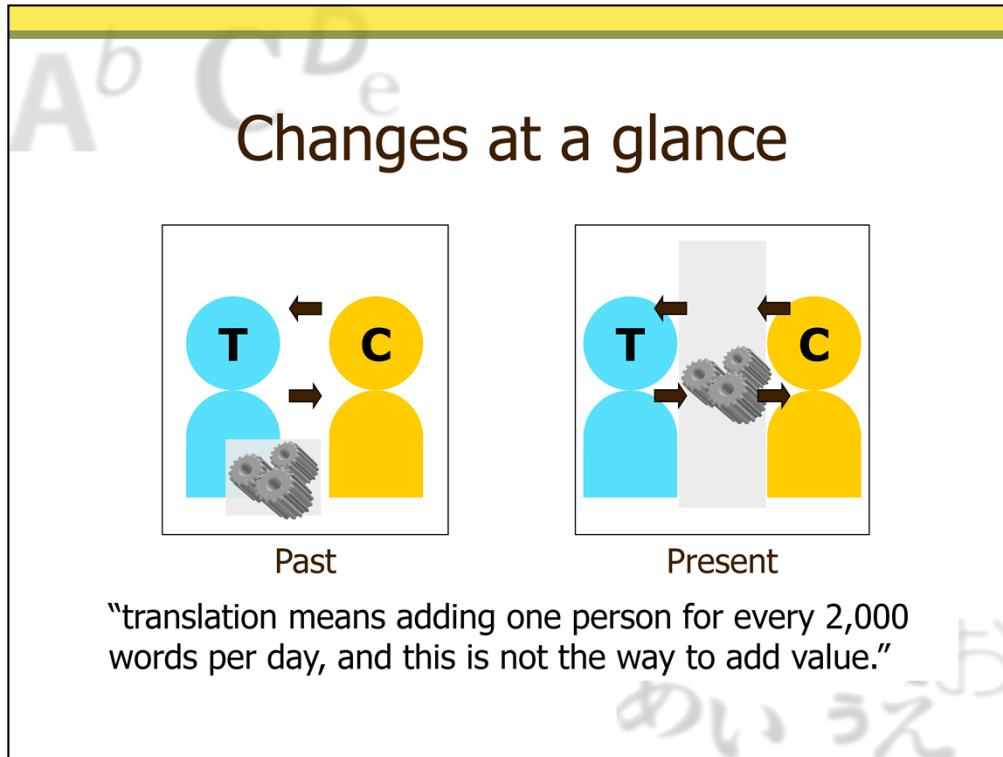
I made this little drawings to represent how all these changes look like in my view.

The blue figure with the letter "T" on her head represents translators. The yellow figure with the letter C represents clients.

The gray box with cogwheels represents translation tools.

The translation tool which was once a nifty little tool box for translator, in the picture on the left, has become a big fat technology layer that lies between translators and their client, as shown on the right.

Also in the picture on the right, notice how I made the tool box taller than people on each side.

It represents how individuals on either side are invisible from each other.

The top and bottom arrows represent how their interaction takes place indirectly, through the technology layer.

Having a little room for flexibility, translator's output became more prone to be homogeneous.

And this contributed to the notion among many LSPs that translators are all the same.

As a Chief Executive Officer of a global service provider company said:<click>

 "translation means adding one person for every 2,000 words per day, and this is not the way to add value."
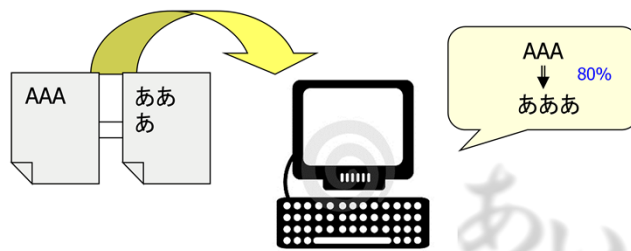
Future?

- Statistical Machine Translation (SMT)
- Convergence of Translation Memory and Machine Translation (MT)
- Machine Translation + post-edit

As for the future trend, I have these three points.

- Statistical Machine Translation

- Convergence of Translation Memory and Machine Translation

- Machine Translation + post-edit

Statistical Machine Translation, or SMT, is a form of Machine Translation that uses statistical model to generate translation.

The statistical model is made by analyzing a vast amount of bilingual text corpus. Bilingual text corpus is a database of source sentences and target sentences.

Computer will be trained to calculate probability of given word, phrase, or sentence to be translated into, based on the parameters derived from the analysis of bilingual text corpus.

For example if the word AAA has the probability of 80% to be translated into ああ あ, then it translate AAA into あああ at 80% confidence.

The more sentences the bilingual text corpus has, the better the statistical model will be, allowing the SMT program to produce more accurate and naturally sounding translation.

The most well-known example of SMT is Google Translate. Google is said to be gathering bilingual text corpus from many documents, by scanning the original version books and their translated version or by crawling websites which have two or more language versions.

Future?

- Statistical Machine Translation (SMT)
- **Convergence of Translation Memory and Machine Translation (MT)**
  - TM = Corpus

Convergence of Translation Memory and Machine Translation. In the previous slide, I said Bilingual text corpus is a database of source sentences and target sentences.

As I mentioned earlier, the Translation Memory is also a database of source sentences and target sentences.

In other words, a translation memory that is accumulated large enough can serve as a bilingual text corpus to make a statistic model for SMT.

In fact, the significant resurgence in interest in machine translation has a lot to do with the fact that many companies have accumulated large amount of translation memory assets from the past translation projects, and they are looking to utilize these assets.

Future?

- Statistical Machine Translation (SMT)
- Convergence of Translation Memory and Machine Translation (MT)
  - TM = Corpus
- **Machine Translation + Post-Edit**
  - **The machine translates, then human edits**
  - **Non-Monetization vs. Monetization Content**

Machine Translation + Post-Edit.

The first two bullets, Statistical Machine Translation, and Convergence of Translation Memory and Machine Translation, all come down to this.

Let the machine translate the document, then have human translators edit the result to make it publishable.

Corporation with needs to communicate vast amount of information to multi-lingual audiences very quickly

while controlling cost are heading to this model.

At JTF conference last December, there was a very interesting presentation in this topic by Reiko Saito,

who is Senior Language Specialist at Oracle Japan.

What I found most interesting is that the model begins with classifying their translatable contents into Non-monetization, which is the contents that does not generate income, and Monetization, which generate income.

Future?

- Machine Translation + Post-Edit
  - Non-Monetization vs. Monetization Content

Troubleshoot, UI, Ref	Paid Support Data	Documentation
MT Only	MT + Light E	MT + Full E
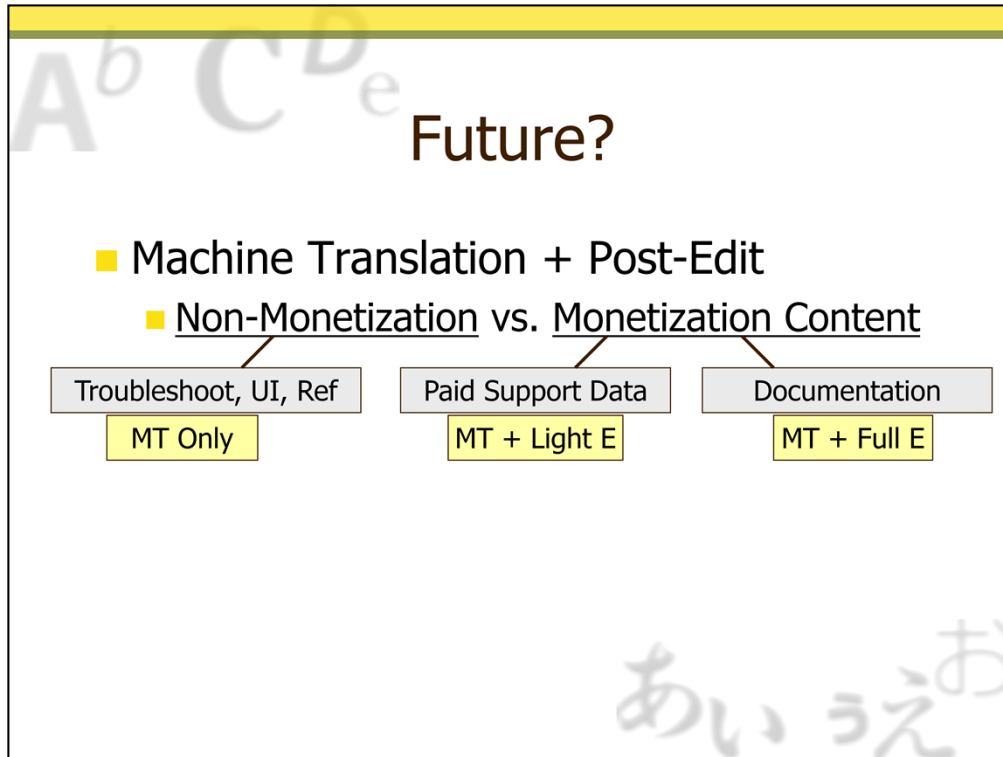
Non-monetization content are vast amount of information that are provided for free. It also consists of non stylistic, simple, and monotonous text. Example includes troubleshooting articles, User Interface and other references provided for free. These content will be translated only with Machine translation.

Monetization Content is further divided into 2 groups. First group requires some level of quality, but it is also cost-sensitive due to large volume. Example includes paid support data.

The second group requires high level of quality, meaning the level of human-translator quality.

It includes documentation provided with fee or important public relation material.

The first group is managed by Machine translation plus Light Edit, and the second group is managed with MT plus Full Edit.

Full Edit is what we know today as standard editing task, which is editing for accuracy and readability.

Future?

- **Machine Translation + Post-Edit**
    - **Non-Monetization vs. Monetization Content**

Troubleshoot, UI, Ref	Paid Support Data	Documentation
MT Only	MT + Light E	MT + Full E
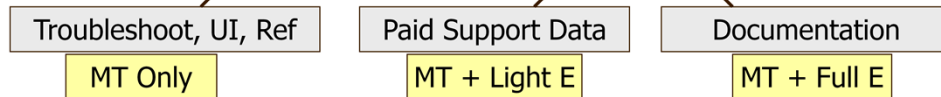
- **Light Edit:** ○ Understandable enough.
   × This is how I would translate.

Extremely tight deadline to set a cap on quality.
→ Work even more faster and cheaper.

Light Edit came as a new idea to me. It is editing only for the accuracy: Accuracy in terms of the information being not wrong, and being able to understood by its intended readers enough for them to achieve the intended action.

"This should be understandable enough." is the expected level. Many translators tend to edit based on "This is how I would translate," but this is not what they are asking.

<click>

In order to discourage translators from editing extensively, the client imposes extremely tight time limits, like matter of few hours per projects, so that translators can spend very limited time on each sentence.

Changes at a glance

Past   Present   Future?
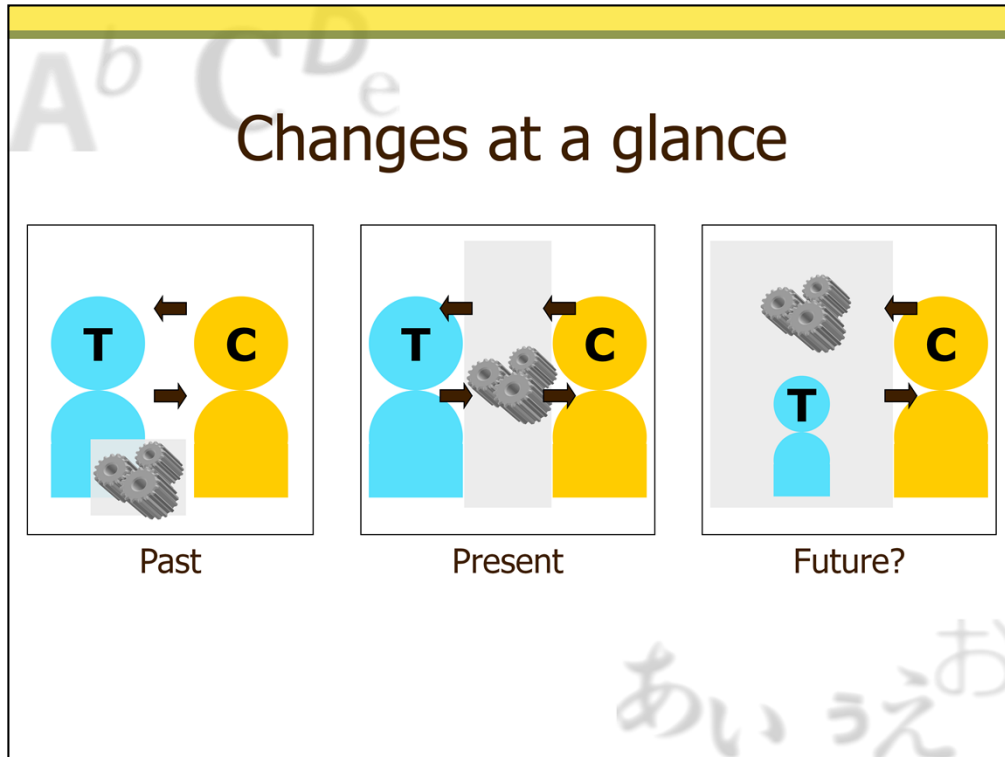
That said, it looks like the Future is heading to this. Translators are reduced to a component of a translation tool.

But for some translators, this Future is already now.

In fact...

CAT/MT Adaptation Models

- Less compatible sub.
- CAT is not common

Localization, Globalization
- Compatible with CAT
- Increase in Volume
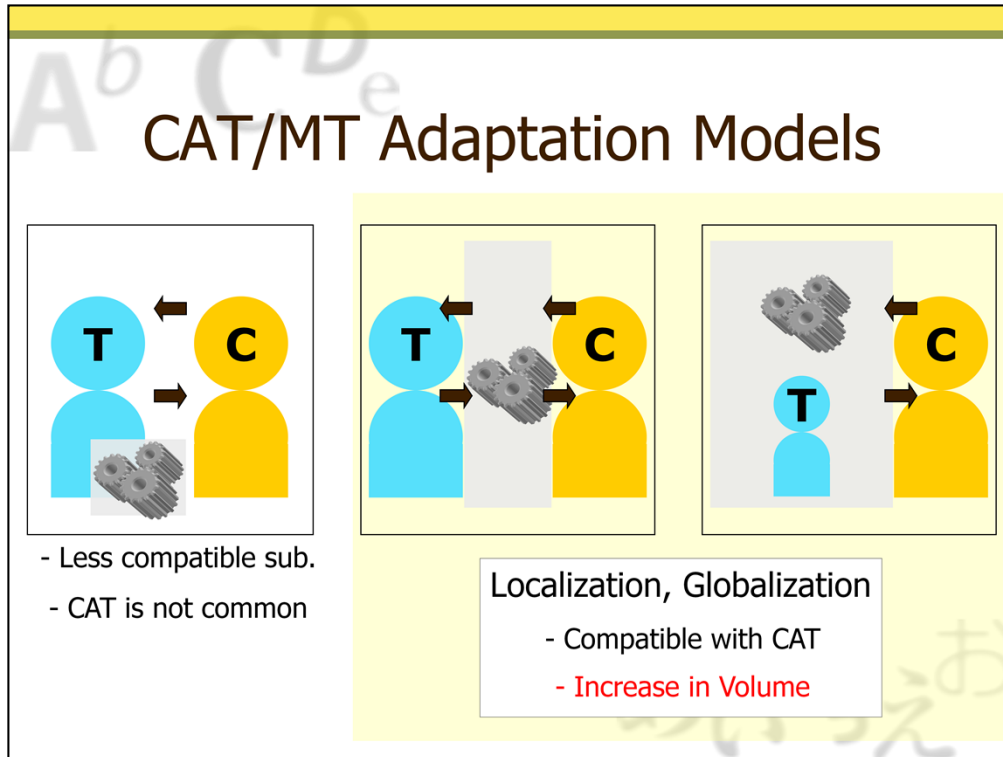
All these models can be all coexisting in present time, each representing a different market segment.

<Click 1>

If you work in the subject field that is less compatible with the Computer Assisted Translation technology, then your market segment may not have adapted the CAT tools yet.

And you may still be working in the model on the left, which I explained earlier as the PAST model.

<Click2>

At present times, the models in the middle and to the right more typically apply to the field of software localization, or globalization of any technology-related contents.

I believe Statistical Machine Translation is also drawing serious attention in Patent translation.

<Click3>

One reason for this, obviously, is the fact that the localization industry is more compatible with computer technology.

But more importantly, <Click4> the key driving force behind this trend is the significant increase in overall translation volumes.

Motivation for CAT/MT

- Internet provides global reach.
  - →Require more translation to engage their multiple target audiences.
- Increased volume, increased cost.
  - →Draws the attention of senior management.
- Can happen to any field of translation.
  - →If the volume and cost reaches the level that justifies it.

In the wake of the Internet, more corporations and organizations are able to expand their reach globally and require more translation to engage their multiple target audiences.

With ever increasing translation volumes, the cost associated with the conventional per-word rate structure has increased significantly; therefore it draw the attention of senior management.

Putting it to this way, the adaptation of the Computer Assisted Translation or even the Machine Translation can happen in any subject area of translation. When the volume and cost reaches the level that justifies it, the clients may start slicing up their translation needs to subdivide into different level of adaptation.

Luckily, we the freelance translators do not need the whole industry, we just need several clients enough to maintain stable income flow.

We may not be able to change where the whole industry is heading, but we just need to steer clear of the area where the task involved in "Translation" is no longer something you enjoy doing everyday. Question is, how we can do that?

Key points to Remember

- Quality of Tool = TM Quality
- TM is only as good as the translator who stores translations in it.
- "Good" Translator can:
  - Produce quality translation
  - Produce quality translation despite the CAT constraints

Going into the subject area where CAT tools are least compatible, is one way.

But for the sake of the argument, lets focus on the case where CAT tools are well adapted, or very likely to be adapted.

First, Quality of Tool equals the Quality of Translation Memory.

In order to obtain high quality result from translation memory tools, it is imperative that the translations stored in the TM is accurate and in good quality.

If the industry is heading toward more hands off approach by utilizing their accumulated translation memories, then the Translation memory quality is more important than ever.

Given the increased competition among LSPs and even among the vendors of translation tools, the quality of TM can affect their competitive edge.

At the end of the day, Translation Memory is only as good as the translator who stores the translations in it.

And the "goodness" of the translator means not just their ability to produce good translation.

It also means their ability to produce good translation despite the range of restrictions imposed on them by the CAT tool's work model.

Translator's CAT skill

- Not established and recognized.
- Hard to demonstrate in trial translation.
- Lack of it does impact quality.
  - Increased job in translation QA
  - "Quality-based Pricing" clause

Know your tools:
- For your own productivity
- A differentiation factor

Translators' specific skill sets to produce quality translation despite CAT constraints is a bit tricky to sell.

For one, this distinction of skills is not quite established and recognized.

And for another, this is not something you can demonstrate in a short passage that agencies give you as trial translation.

But I suppose that organizations do feel the pinch associated with lack of such skill.

It is evident in the increased job posting for translation quality assurance.

Recently I came across an agency who has "Quality-based Pricing" clause in its agreement.

It asks translators to agree that their work will remain at least at the quality level of their translation test.

If the quality continued to fail without improvement, they will reduce the translator's per-word or per-hour rate.

I suppose this agency must have experienced some quality issues, and wondered why some translators who do well in the trial and then not do well in daily translation jobs.

<Click>

With these points considered, it is important for translators to know their tools. In terms of what it does to you, both in good and bad ways.

It is good for your own productivity. And moreover, it may serve you as a differentiation

factor.

TM Tool as a Differentiation Factor for Translators
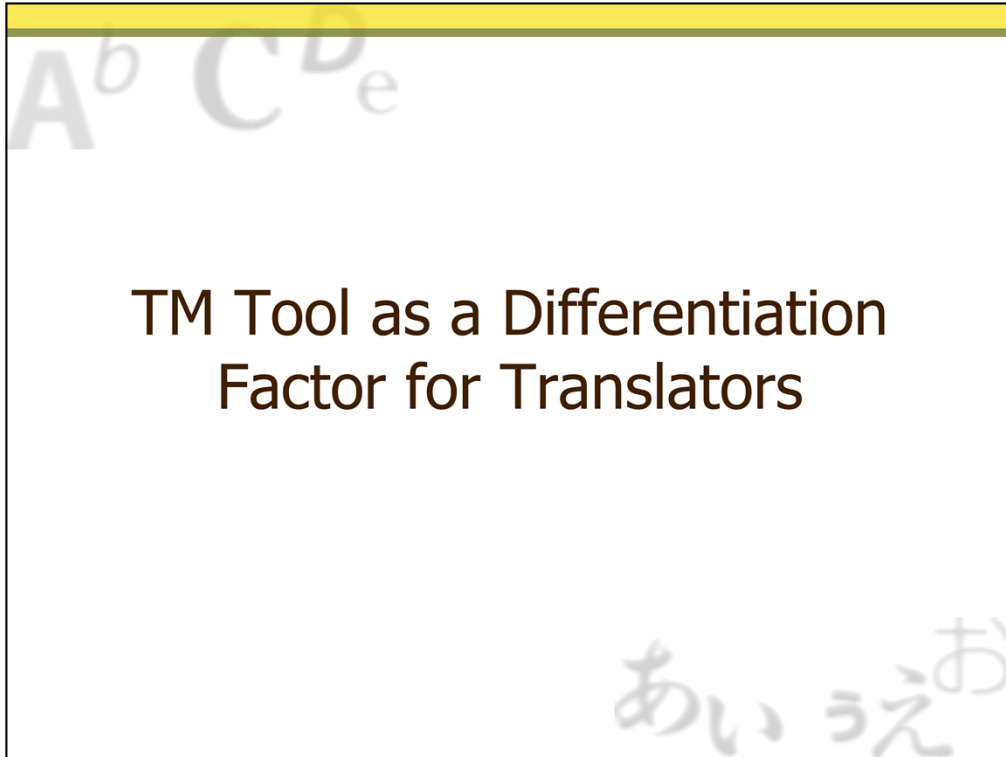
Now, I would like to discuss Translation Memory Tool as a differentiation factor for translators.

Shortfalls of TM tools

- Tool Translationese

- Primacy of Existing Translations

- Decentralization of Translation Assets

I would like to start with identifying common Shortfalls of Translation Memory tools.

I identified three key areas:

Tool Translationese, Primacy of Existing Translations, and Decentralization of Translation Assets.

Let's look into these points, one by one.

Tool Translationese

- Not necessarily incorrect but awkward translation.

- Caused by the tool's mechanical restriction imposed on how translators translate.

- Reinforced by a business model that emphasizes TM recyclability and allows little flexibility to translators.

Tool Translationese is...

not necessarily incorrect but awkward translation.

It is caused by the tool's mechanical restriction imposed on how translators translate, and

It is reinforced by a business model that emphasizes TM recyclability and allows little flexibility to translators.

Tool Translationese - 1

- Segmentation rule

Supported languages include: Java, HTML/JavaScript, XML, and SQL.

T1: 対応言語は以下を含みます: Java、 HTML/JavaScript、XML、SQL。

T2: Java、HTML/JavaScript、XML、SQL などの言語に対応しています。

The first case of tool translationese is when translator has to compromise her rendition in consideration for segmentation rule.

A typical example of this would be a sentence that is divided by a colon in the middle.

Please take a moment to read the sample text shown on the screen.

[Wait for audience to finish reading]

T1 and T2 are both translation for this source text.

Which one do you think is better translation, in terms of the style which native Japanese speaker would write this sentence in Japanese from scratch.

I think it's T2.

Tool Translationese - 1

- Segmentation rule

    Supported languages include: Java, HTML/JavaScript, XML, and SQL.

    T1: 対応言語は以下を含みます: Java、HTML/JavaScript、XML、SQL。

    T2: Java、HTML/JavaScript、XML、SQL などの言語に対応しています。

This source text appears to be a single sentence, but it is actually consists of 2 separate segments.

Most CAT tools in default setting, treat a colon as a stop character to end the segment.

So, in order to translate this sentence as T2, the translator must either:
 -connect these 2 segments into 1 segment, or
 -start with entering the translation of the latter half of this sentence <Click1> into the first segment <Click2>,

         and then enter the translation of the first half of the sentence <Click3> into the second segment <Click4>

Connecting 2 segments into one involves modifying a preset property of translation memory.

While the tools most often offer a command to allow this particular action, translators are often instructed by their client not to connect or divide existing segment, in order to maximize the recyclability of the translation memory.

Entering translation for other source segment may allow you to keep the segmentation rule intact,

but it also plants a seed of mistranslation if any of these segments are recycled in a different context.

Both cases involve a risk of disrupting the translation memory, which is an asset that belongs to your client.

To avoid this risk, the translator may choose to translate as T1. <Click5>

Although it is less ideal, it can still carry the intended meaning.

Tool Translationese - 2

- **1 vs. 1 Pair Restriction**
  - 1 version of TR per source text.
  - Troublesome if variations of TR needed for the exact same original.
    - Yes = "はい" or "有" or "可"
    - Index marker text

Create New → 新規作成[しんきさくせい]
Create New → 新規作成
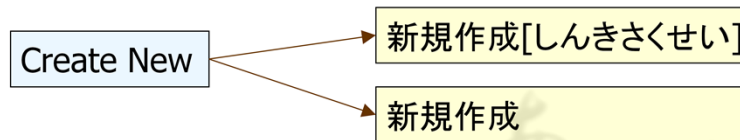
The second case happens due to 1-to-1 pair restriction typically imposed by most translation tools.

Because translation memory saves the translated segment as 1-to-1 pair with the source segment, only 1 version of translation is allowed for any given source segment.

This poses trouble if a same original segment needs to be translated differently depending on the context.

For example, a segment as simple as "Yes" may require this many variations. This a real example which is taken from actual assignment I've done in the past.

Another typical example is index marker text. Index marker text is a kind of text used by Desktop Publishing programs to generate alphabetically-sorted index.

In English to Japanese translation, it is conventional to add *yomigana* to the translation of Index Marker Text so that non-phonetic characters like Kanji can be sorted appropriately.

 <Click>

If the index marker text uses exact same English phrase as a text in manual content, which is often the case for section titles or table headings, you will need 2 different translation for this source text. One with *yomigana* and one without.

Tool Translationese - 3

- **Peephole Translation (Heyn 1998, 135)**
  - **Definition:**
    [a phenomenon in which] translators deliberately formulate texts in such a way as to make them more easily recyclable from a TM. ...This may result in text that is less cohesive and consequently less readable. (Bowker, 2002)

  - **Discourages common translation techniques**
    - Omit or replace pronouns
    - Add conjunctions

The third case is a phenomenon described as 'peep-hole translation.'

It is a phenomenon in which translators deliberately formulate texts in such a way as to make them more easily recyclable from a translation memory. ...This may result in text that is less cohesive and consequently less readable.

More specifically, translators are discouraged from applying common translation techniques such as:

 - omitting pronouns or replacing them with proper noun or other descriptive phrases; and

 - Adding conjunctions.

Both are known techniques in E to J translation to improve cohesion between sentences or paragraphs; however,

omitting something from original text, or adding something that wasn't in the original involves a relative decision. And the result may change depending on the context.

Translators working with Translation Memory tools may refrain from making such decisions in consideration for maximum recyclability of the translation memory.

Tool Translationese - 4

- Sentence Salad
  - Correction of incoherent sentences.
  - Aggravated when the TM comprises a variety of different texts that have been translated by different translators.
  - Result from "Peephole Translation."
  - Because translators are forced to work on one segment at a time.

The next case is a phenomenon described as Sentence Salad.

Sentence Salad is:

•a correction of incoherent sentences.

•It is aggravated when the TM comprises a variety of different texts that have been translated by different translators.

•It can be a result from "Peephole Translation."

•It is also attributed to the fact that the translation tool's interface forces translators to work on one segment at a time.

Sentence Salad: Example

With production costs and creative integrity at stake, compromising on digital video tools is simply not an option. You need to know that you got the shot, and the best time to know that is on site, when everyone and everything is still in place. The XYZ and XYZ Plus portable server and workstations have the performance, reliability and flexibility to deliver that confidence.

Tr1: 制作コストと創造的完璧さがかかわる以上、デジタル映像ツールについての妥協は許されません。望みのショットが撮れたことを確認することが必要であり、そのために最良なのは、撮影現場で、関係者や機材すべてがそのままである時なのです。XYZ と XYZ Plus ポータブル・サーバおよびワークステーションは、その確かさを実現するパフォーマンス、信頼性、そして柔軟性を備えています。

This is an example of sentence salad. This text is taken from actual document I found in the internet.

Please take a moment to read the sample text shown on the screen.

[Wait for audience to finish reading]

Do you sense the awkwardness? Just by looking at this text, I can tell this text has been translated by someone using translation memory tool.

Each sentence is not wrong grammatically, and it does seem to carry the original meaning.

But as a whole paragraph, the relationship between sentences, or cohesion, is very weak.

Sentence Salad: Example

① With production costs and creative integrity at stake, compromising on digital video tools is simply not an option. You need to know that you got the shot, and the best time to know that is on site, when everyone and everything is still in place. The XYZ and XYZ Plus portable server and workstations have the performance, reliability and flexibility to deliver that confidence.

Tr1: 制作コストと創造的完璧さがかかわる以上、デジタル映像ツールについての妥協は許されません。望みのショットが撮れたことを確認することが必要であり、そのために最良なのは、撮影現場で、関係者や機材すべてがそのままである時なのです。XYZ と XYZ Plus ポータブル・サーバおよびワークステーションは、その確かさを実現するパフォーマンス、信頼性、そして柔軟性を備えています。

This kind of awkwardness is most often caused by translator's working on one sentence at a time, as it presented by the translation tool.

Like this.

<Click>

Sentence Salad: Example

With production costs and creative integrity at stake, compromising on digital video tools is simply not an option. You need to know that you got the shot, and the best time to know that is on site, when everyone and everything is still in place. The XYZ and XYZ Plus portable server and workstations have the performance, reliability and flexibility to deliver that confidence.

②

Tr1: 制作コストと創造的完璧さがかかわる以上、デジタル映像ツールについての妥協は許されません。望みのショットが撮れたことを確認することが必要であり、そのために最良なのは、撮影現場で、関係者や機材すべてがそのままである時なのです。XYZ と XYZ Plus ポータブル・サーバおよびワークステーションは、その確かさを実現するパフォーマンス、信頼性、そして柔軟性を備えています。

Sentence Salad: Example

With production costs and creative integrity at stake, compromising on digital video tools is simply not an option. You need to know that you got the shot, and the best time to know that is on site, when everyone and everything is still in place. The XYZ and XYZ Plus portable server ③ and workstations have the performance, reliability and flexibility to deliver that confidence.

Tr1: 制作コストと創造的完璧さがかかわる以上、デジタル映像ツールについての妥協は許されません。望みのショットが撮れたことを確認することが必要であり、そのために最良なのは、撮影現場で、関係者や機材すべてがそのままである時なのです。XYZ と XYZ Plus ポータブル・サーバおよびワークステーションは、その確かさを実現するパフォーマンス、信頼性、そして柔軟性を備えています。

Sentence Salad: Example

The main idea for this entire paragraph is in the last sentence.

"The XYZ and XYZ Plus portable server and workstations have the performance, reliability and flexibility to deliver that confidence."

The first two sentences are supporting ideas.

<Click1>

In the first sentence,

<Click2>

"compromising on digital video tools is simply not an option" is the key supporting idea.

<Click3>

In the second sentence,

<Click4>

"the best time to know that is on site" is the key supporting idea.

Taking these context information into account will make the translation more readable and natural.

Sentence Salad: Corrected

With production costs and creative integrity at stake, compromising on digital video tools is simply not an option. You need to know that you got the shot, and the best time to know that is on site, when everyone and everything is still in place. The XYZ and XYZ Plus portable server and workstations have the performance, reliability and flexibility to deliver that confidence.

Tr2: デジタル映像ツールの性能は制作費や作品の完成度に直接影響を及ぼすものであり、そこに妥協は許されません。望みのショットが撮れたかを知る必要があることはもちろんですが、スタッフや機材が揃っている撮影現場に居ながら映像を確認できることが理想です。XYZ および XYZ Plus ポータブルサーバーとワークステーションはその理想を実現する性能、信頼性、そして柔軟性を備えています。

That said, this is a corrected version of the translation.

If the translator works with awareness of the sentence salad issues, this level of translation can be achieved even if translation tools are used.

Tool Translationese - 5

- Output along with reading the original
  - Place information in the same order as the original.
  - Disregard the context flow appropriate for the target language.

The last example of tool translationese occurs when translator output her translation along with reading the original.

As she reads the sentence left to right, as soon as a unit of meaning completes, she transfer it to target language,

then moves on to the next meaning unit.

As a result, the information is placed in the same order as the original, disregarding the context flow that is appropriate for the target language.

It is very much like sentence salad phenomenon, but this time it can happen within a single sentence.

Output as TR read the original: Example

Main Idea

This simplifies account administration, allowing you to apply a filter to multiple properties at once, or to see all of the subdomains you're tracking in separate properties all at once.

Tr1: これによりアカウントの管理が単純化され、フィルタを複数のプロパティにまとめて適用することができます。また、個別のプロパティに含まれるトラッキング対象のすべてのサブドメインをまとめて表示することもできます。

Wrong emphasis

This is an example of resulting translation. Please take a moment to read both the original and the translation.

[Wait till audience reads the text]

The main idea of this sentence is, "This simplifies account administration".<Click1>

This is the only independent clause of this sentence.

Everything after the comma <Click2> are subordinating clauses supporting the main idea.

As a result of translating each phrases in the order of its appearance, however, <Click 3>

resulting translation has emphasis on a wrong place.

Output as TR read the original: Correction

This simplifies account administration, allowing you to apply a filter to multiple properties at once, or to see all of the subdomains you're tracking in separate properties all at once.

Tr2: そうすることで、複数のプロパティにまとめてフィルタを適用したり、別々のプロパティの中でトラッキングしているサブドメインをまとめて表示したりといった操作が可能になり、アカウントの管理がいっそう簡潔になります。

Considering this, better translation for this original would be something like this.

Primacy of Existing Translations

- Translators readily accept existing translation suggested by TM.

- Fuzzy M. that is awkward but not wrong
  → Follow the same style rather than re-writing it.

- TM continues to accumulate "less than ideal" translation, reinforcing it as prevailing style.

Primacy of Existing Translations is social and political aspect of translation tool issues.

•Translators readily accept existing translation suggested by TM.

•When Fuzzy Match translation suggested by the translation memory is awkward but not wrong, translators are most likely to accept it and follow the same style rather than re-writing it.

•As a result, the translation memory continues to accumulate "less than ideal" translation, reinforcing it as prevailing style.

Primacy of Existing Translations

- Reason: No merit to correct it.
  - Only paid for translating the differences.

Primacy of Existing Translation exists because there is no merit for translators to correct existing translations.

After all, we are only paid for translating the differences.

<Click>

Primacy of Existing Translations

- Reason: No merit to correct it.
  - Only paid for translating the differences.
  - Risk of "consistency error."

And modifying existing translation may have a risk of being counted as a consistency error by your evaluator.

"Consistency error" is normally counted if translator deviate from the style guide and glossary.

Deviating from the style generally accepted in other parts of document may also counted as a consistency error.

Primacy of Existing Translations

- Reason: No merit to correct it.
  - Only paid for translating the differences.
  - Risk of "consistency error."
  - Risk of reduced TM recyclability, which may not be welcomed by the client.

It also risks reduced TM recyclability, which may not be welcomed by the client.

Primacy of Existing Translations

- **Reason: No merit to correct it.**
  - Only paid for translating the differences.
  - Risk of "consistency error."
  - Risk of reduced TM recyclability, which may not be welcomed by the client.
  - **Existing translation = Low-risk option.**

And most of all, existing translation is always a low risk option for translators.

Primacy of Existing Translations

- **Reason: No merit to correct it.**
  - Only paid for translating the differences.
  - Risk of "consistency error."
  - Risk of reduced TM recyclability, which may not be welcomed by the client.
  - **Existing translation = Low-risk option.**
    - TM's authority for being provided by the client

For subcontracted translators, the translation memory provided by their client along with the job itself poses certain authority.

Primacy of Existing Translations

- **Reason: No merit to correct it.**
  - Only paid for translating the differences.
  - Risk of "consistency error."
  - Risk of reduced TM recyclability, which may not be welcomed by the client.
  - **Existing translation = Low-risk option.**
    - TM's authority for being provided by the client
    - Translated/edited by someone already paid for → You have someone else to blame.

And of course, the translations contained in that translation memory have been previously translated and then edited by someone else. And supposedly they have already been approved and paid for those work.

So, if anything goes wrong, you may have someone else to blame.

Decentralization of TR Assets

- Multiple versions: TM, Termbase, and Translated docs can be modified independently.
- Separate glossaries, style guide, and other reference materials to look up.
- No single authority to resolve the inconsistency.

Decentralization of Translation Assets.

Although they work together under integrated interface, the translation memory, termbase, and translated document are independent files under separate applications. So, anything translated using translation memory tool will end up being stored in at least 3 separate files.

And because each of these files can be edited and modified independently, it is often the case that each of these files have different versions of translation.

Partly because of this issue, a translation project typically comes with:

-a separate glossaries in spreadsheet format

-style guides and

-other reference materials

and translators are asked to comply with each of them as they work on the project. That is a lot of document to go back and forth.

Sometime you even have more than one glossary, each with different purpose.

It takes lot of time and memorization to go through all there materials and stay familiar with them, especially for a freelance translator working with multiple projects.

And to make the matters worse, it is often the case that there is no single authority who can make decision to resolve these inconsistency.

It is because there is some imbalance of information available to each involved party, that is, translators, agencies, and the end client.

Quality Issue for Client

- Translationese issues are not fully manageable by <u>Quantitative QA</u>*.
- Accumulation of less-than-ideal translation in TM sets a cap on overall quality.
- Identifying and correcting problems in a massive TM is a daunting task.

* QA by scoring based on predefined error categories

And these translates into quality issues for client.

From the client's point of view, they lead to the following issues:

First, Translationese issues are not fully manageable by Quantitative Quality Assurance approach.

Quantitative Quality Assurance is an approach increasingly common in localization field.

It evaluates translation by scores based on predefined error categories, such as mistranslation, typographic error, consistency error, and terminology error.

Second, the primacy of existing translation will result in accumulation of less-than-ideal translation in TM,

which will set a cap on overall quality.

And identifying and correcting problems in a massive TM is a daunting task. Given the decentralization of entire translation assets, going back to what has already been translated to make some adjustments could trigger all sorts of consequences.

So, it only make sense to do it right at the first time.

Turn Them into Added Value

- Be aware of translationese cases and able to work around them.
- Effective use of translator's note: Leave paper trail of issues.
  - Bring attention to undetected issues.
  - Protection for yourself if things go wrong.
- Provide traceable record that may help client to identify problem area later on.

So why don't we turn them into our added value?

Translator with right skills and experience can be aware of translationese cases and able to work around them.

We can use translator's note more effectively, and that will leave paper trail of issues.

It will bring attention to undetected issues before resolving such issue gets too complicated.

Even if it did not trigger any immediate action, it still can provide protection for yourself in case things did fall apart.

And, if you continue to provide effective translator's notes, that will provide traceable record that may help client to identify problem area later on.

And that will be one good reason for the client to prefer working with you.

Do's and Don'ts

- Do...
  - Take ownership of your work.
    - Be professional.
    - Always translate at your best.
    - If TM is in your way, report it.
  - Be objective and effective communicator.
    - Skills to edit other's work or own work.
    - Skills to be edited by others.
    - Be accountable.

With that in mind, here is the Do's and Don'ts for individual translators.

Do...

    Take ownership of your work.

        Be professional.

        Always translate at your best.

        If TM is in your way, report it.

    Be objective and effective communicator.

        Skills to edit other's work or own work.

        Skills to be edited by others.

        Be accountable.

Do's and Don'ts

- Don't...
  - Blame everything on TM or Termbase.
    - "Because TM/Termbase said so" may work but will hurt you in the long run.
  - Review other's work with hostility.
    - It WILL backfire at the rebuttal stage.
  - Take reviewer's comments personal.
    - Opportunity to demonstrate your objectiveness.

Don't...

•Blame everything on TM or Termbase.

   "Because TM/Termbase said so" may work but will hurt you in the long run.

When i review, I see a lot of people using this excuse. If you voluntarily give up your active role in the creative process of translation, it only make it easier for LSPs to put you aside for more passive role in the "Machine translation plus post-edit" approach, which I showed you earlier.

•Review other's work with hostility.

Occasionally I come across a reviewer who wrongly feel empowered by their task.

In many case, the review process is followed by a rebuttal stage to allow the original translator to give feedback.

If you make baseless accusations in attempt to fend off competition, they may blow up in your face at the rebuttal stage.

Review job is not a place for smear campaign. It is for identifying issues and communicating them objectively and effectively.

The review result may also become a part of important translation asset for your client.

Your performance as a reviewer is evaluated based on your ability to review. Right way to compete is to provide useful result.

•Take reviewer's comments personal.

I also review other people's work. And no matter how I take extra care to reason my corrections politely and objectively, some translators return emotional feedback.

We are both on the same side, with a common goal of achieving best possible translation for the project.

This is your opportunity to demonstrate your objectiveness, your ability to understand the issue and resolve it quickly.

Everybody makes mistakes. Being able to acknowledge and correct them effortlessly is very practical and valuable skill.

Appeal to the Industry

- Bring attention to the translator's skill required to produce good translation despite the restriction of CAT tools.
- Make such skills measurable and recognizable.
  - Certification
  - Handbook

By a translator's organization.
Not by certain CAT tool vendors or translation companies.

To appeal human translator's value to the translation industry, translators should:

•Bring attention to the translator's skill required to produce good translation despite the restriction of translation tools.

and

•Make such skills measurable and recognizable.

I wonder if we can do this by offering a certification or adding a new category to existing certification.

OR, prepare a handbook for working with computer assisted translation tools.

<<Click>>

It is very important that these initiatives are taken by a translator's organization.

Not by certain translation tool vendors or translation companies.

Tool vendor's focus is locking in their users to secure market share, and translation companies' are more interested in manageability of translation memory.

Both does not represent our interest.

72

Conclusion

- The question is no longer about whether to use translation tool or not. It is about how to use it.
- Translator is still a single largest contributor of quality translation memories.
- Voice our perspective to keep translation something worth doing.

In conclusion...

•The question is no longer about whether to use translation tool or not. It is about how to use it.

•Translator is still a single largest contributor of quality translation memories.

•Voice our perspective to keep translation something worth doing.

And this presentation was my attempt to voice my perspective, to start.

References

- Computer-aided translation technology: A practical introduction.
  Bowker, L. (2002). Ottawa: University of Ottawa Press.
- Fuzzy matching (Wikipedia)
  http://en.wikipedia.org/wiki/Fuzzy_matching
- How Google Translate Works
  http://hubpages.com/hub/How-Google-Translate-Works
- LISA Best Practice Guide: Quality Assurance – The Client Perspective
  Localization Industry Standards Association
- On Toury's laws of how translators translate
  Pym, A. (2008). http://www.tinet.org/~apym/on-line/translation/2007_toury_laws.pdf
- Statistical machine translation (Wikipedia)
  http://en.wikipedia.org/wiki/Statistical_machine_translation
- Translation memory (Wikipedia)
  http://en.wikipedia.org/wiki/Translation_memory
- Productivity vs. Quality: A pilot study on the impact of translation memory systems.
  Bowker, L. (2005). Localisation Focus, 4 (1): 13-20.
- ポストエディット入門2010（Introduction to Post Edit 2010）
  Reiko Saito, Presentation at JTF Conference, Dec. 2010

ありがとうございました
Thank You

naoko@japaneseit.net