



It takes community and technologies
to make documents translation work
more fun & productive

Noriaki Fukuyasu

VP of Japan Operations, The Linux Foundation



Key topics of this talk

- › How technologies can solve the issue with translation?
- › What tools we can take advantage of?
- › How can we make it fun and productive?

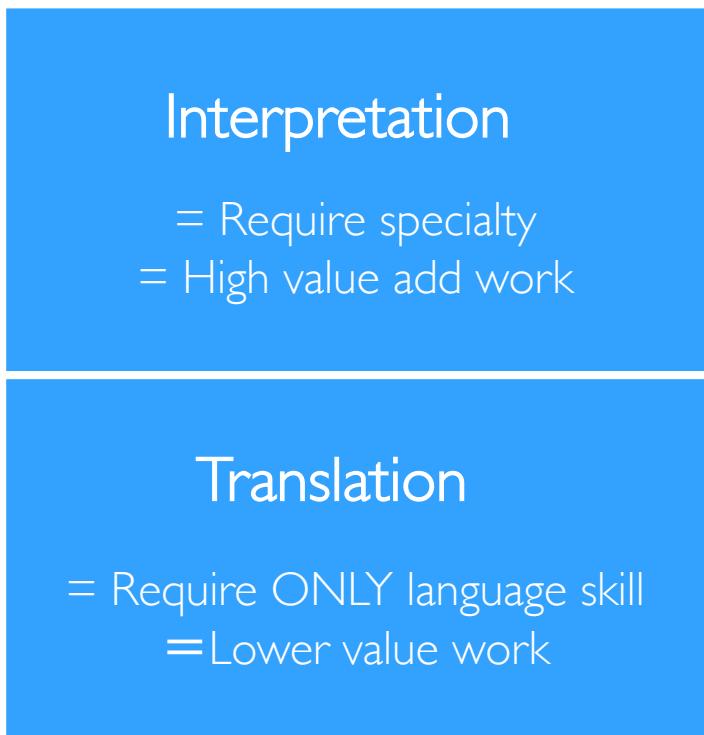
A Big Headache for local open source communities ...

- › The impact of "language barrier" to the industries in non-English speaking countries is HUGE!
(This is particularly true in Japan)
- › Without translation, the new technologies originated overseas will not be spread in the non-English speaking countries (again, like Japan)
- › It is impossible to make ALL Japanese engineers to fluently read English, but it is possible to localize more documents a lot faster by taking advantage of technologies.

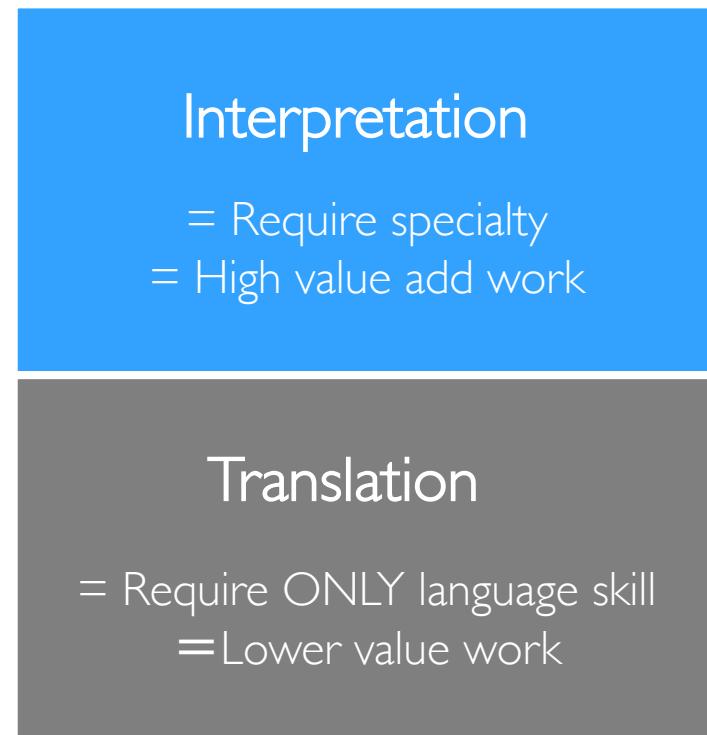
How technologies can potentially solve the issue?

- what is the nature of the work?
- what is the area the technology can potentially intervene?

TODAY = Mainly human work



With Technologies = Mainly by machine



Future of translation business model

- Will the business model change happen just like open source?

OSS Business TODAY



OSS =
Non-Core Competence

Translation Business Future



Translation tools
Machine Translation Engine
=Non-Core Competence
=No need to re-invent wheels

We need an open and common platform for open source communities!

- This should make our translation work a lot easier!
- This will also make our translation work more fun!

OSS Business Today



OSS =
Non-Core Competence

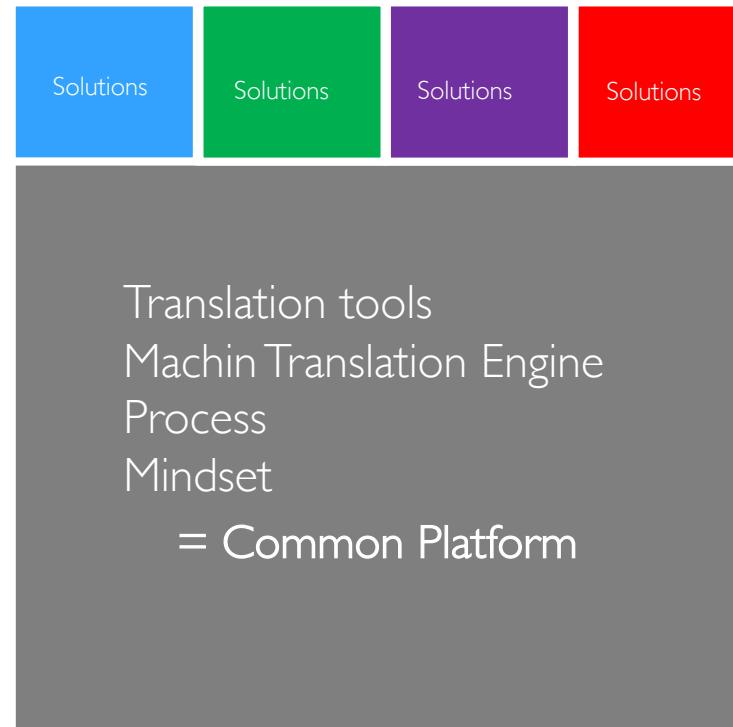
Translation Business Future



Translation tools
Machine Translation Engine=Non-
Core Competence
=No need to re-invent wheels

No Open/Common Platform
(tools, process etc.)

Let's build open and common platform TOGETHER!



What we need?

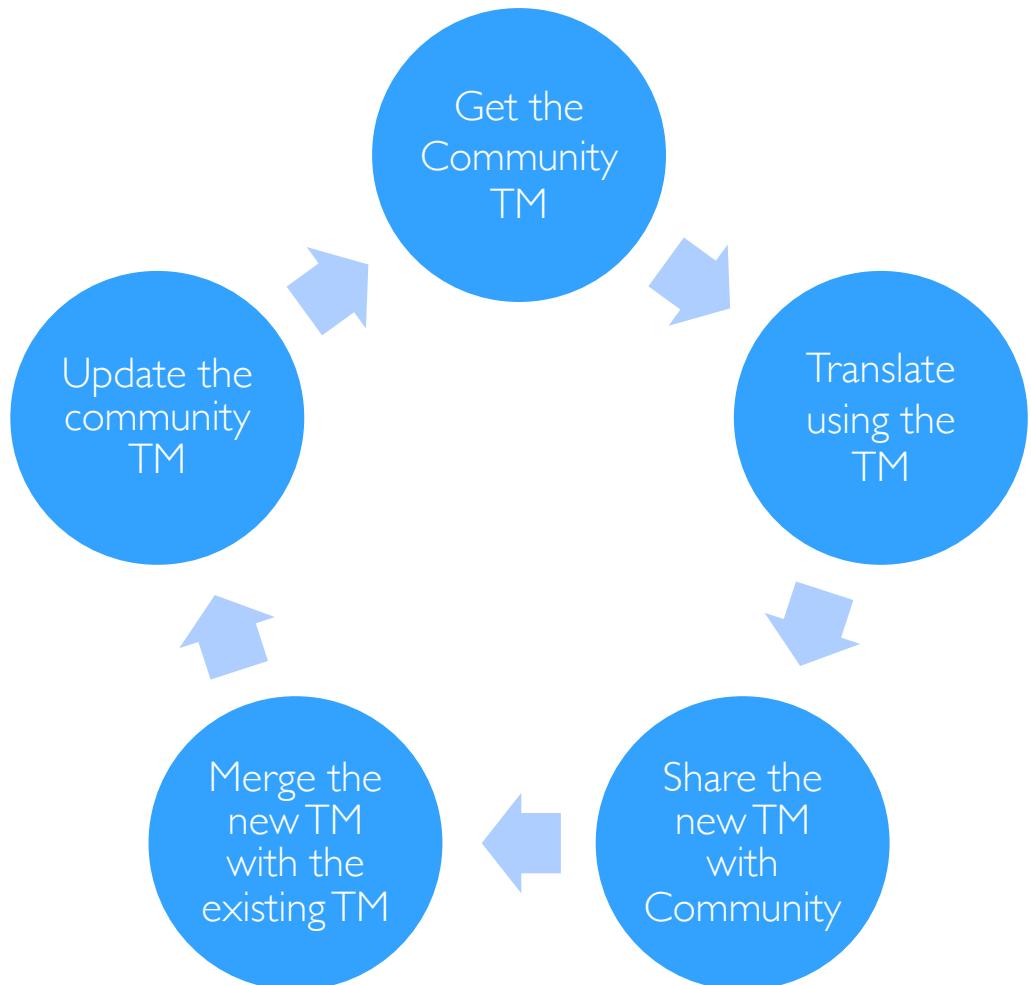
1. Everyone use tools to translate and **SHARE THE TRANSLATION MEMORY!**
2. Train **THE MACHINE TRANSLATION ENGINE** optimized for open source industry.
3. Build a **MINDSET and PROCESS** that the translation will have to be done by the community (never do it alone!) and share the result

What is TRANSLATION MEMORY and why it matters?

- › Translation Memory (TM) is a database that stores “segments”, which can be sentences, paragraphs or sentence-like units (headings, titles or elements in a list) that have previously been translated, in order to aid human translators. (source: [Wikipedia](#))
- › There is an industry standard file format called [Translation Memory eXchange \(TMX\)](#) that will enable people to share and use the memories regardless of the tools they might use.
- › The benefits of using translation memories are:
 - › Avoiding from translating the same sentences/words/sentence-like units over and over
 - › Specific terminologies will be translated into the same local language.
 - › Set translation rules by the community

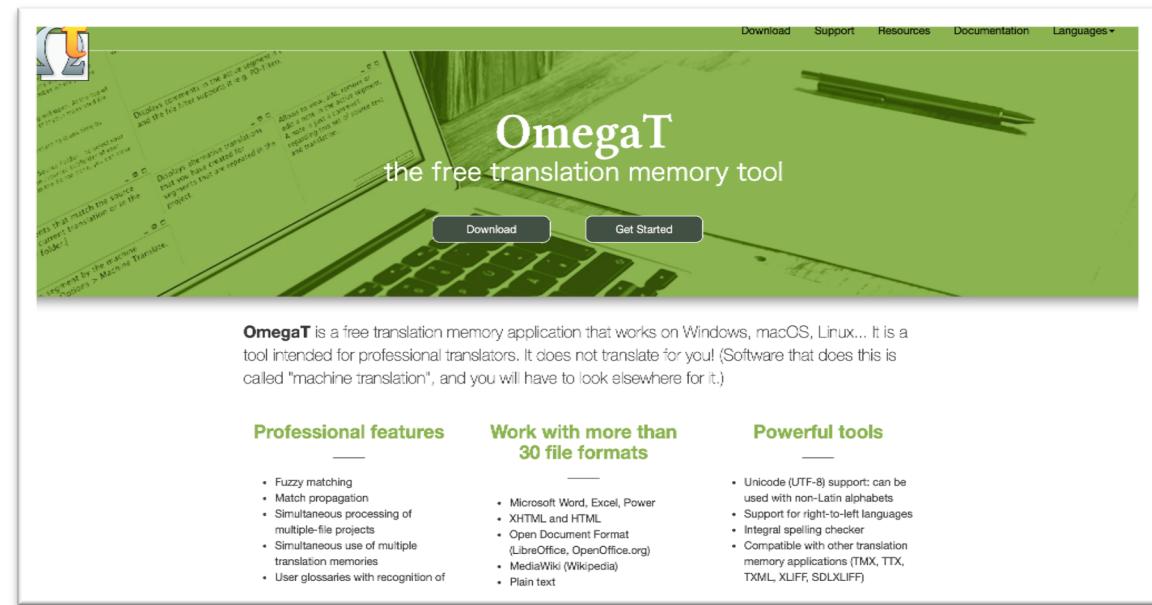
```
2 <tmx version="1.4"><header creationtool="unknown",creationtoolversion="unknown",segtype="paragraph",o-tmfc="unknown",s  
3 clang="en",adminlang="ja",datatype="text"></header><body>\n4 <tu>\n5 <tuv xml:lang="en"><seg> Blockchain and Distributed Ledger Technology</seg></tuv>\n6 <tuv xml:lang="ja"><seg>ブロックチェーンと分散台帳テクノロジー</seg></tuv>\n7 </tu>\n8 <tuv xml:lang="en"><seg>By the end of this chapter, you should be able to:</seg></tuv>\n9 <tuv xml:lang="ja"><seg>本章の終わりまでに、次のことができるようになります。</seg></tuv>\n0 </tu>\n1 <tuv>\n2 <tuv xml:lang="en"><seg>Understand what centralized, distributed, and decentralized network architectures are.</seg></tuv>\n3 <tuv xml:lang="ja"><seg>中央集権型、分散型、および非中央集権型ネットワークアーキテクチャについて理解する。</seg></tuv>\n4 </tu>\n5 <tuv>\n6 <tuv xml:lang="en"><seg>Discuss the problems that blockchain was designed to solve.</seg></tuv>\n7 <tuv xml:lang="ja"><seg>ブロックチェーンが解決するように設計された問題について説明する。</seg></tuv>\n8 </tu>\n9 <tuv>\n0 <tuv xml:lang="en"><seg>Explain differences between permissioned and permissionless blockchains.</seg></tuv>\n1 <tuv xml:lang="ja"><seg>パーミッションドブロックチェーンとパーミッションレスブロックチェーンの違いを説明する。</seg></tuv>\n2 </tu>\n3 <tuv>\n4 <tuv xml:lang="en"><seg>Understand why Hyperledger Fabric was developed.</seg></tuv>\n5 <tuv xml:lang="ja"><seg>Hyperledger Fabricが開発された理由を理解する。</seg></tuv>\n
```

What to do with Translation Memories ?



Available Tools: OmegaT

- › An open source translation tool.
- › Compatible with a variety of translation memory formats (TMX, TTX, TXML, XLIFF, SDLXLIFF)
- › Able to translate over 30 document formats such as Microsoft Word, Excel, Powerpoint, XHTML, HTML, Open Document Format, MediaWiki (Wikipedia) and plain texts.
- › Have interface with popular machine translation engines such as Google, DeepL etc



The screenshot shows the official website for OmegaT. At the top, there's a navigation bar with links for Download, Support, Resources, Documentation, and Languages. Below the navigation is a large banner featuring a green background with a keyboard, a pen, and a small screenshot of the software interface. The text "OmegaT" is prominently displayed in white, followed by "the free translation memory tool". Below the banner is a brief description: "OmegaT is a free translation memory application that works on Windows, macOS, Linux... It is a tool intended for professional translators. It does not translate for you! (Software that does this is called "machine translation", and you will have to look elsewhere for it.)". Underneath this, there are three sections: "Professional features", "Work with more than 30 file formats", and "Powerful tools", each listing specific features or formats.

Professional features

- Fuzzy matching
- Match propagation
- Simultaneous processing of multiple-file projects
- Simultaneous use of multiple translation memories
- User glossaries with recognition of

Work with more than 30 file formats

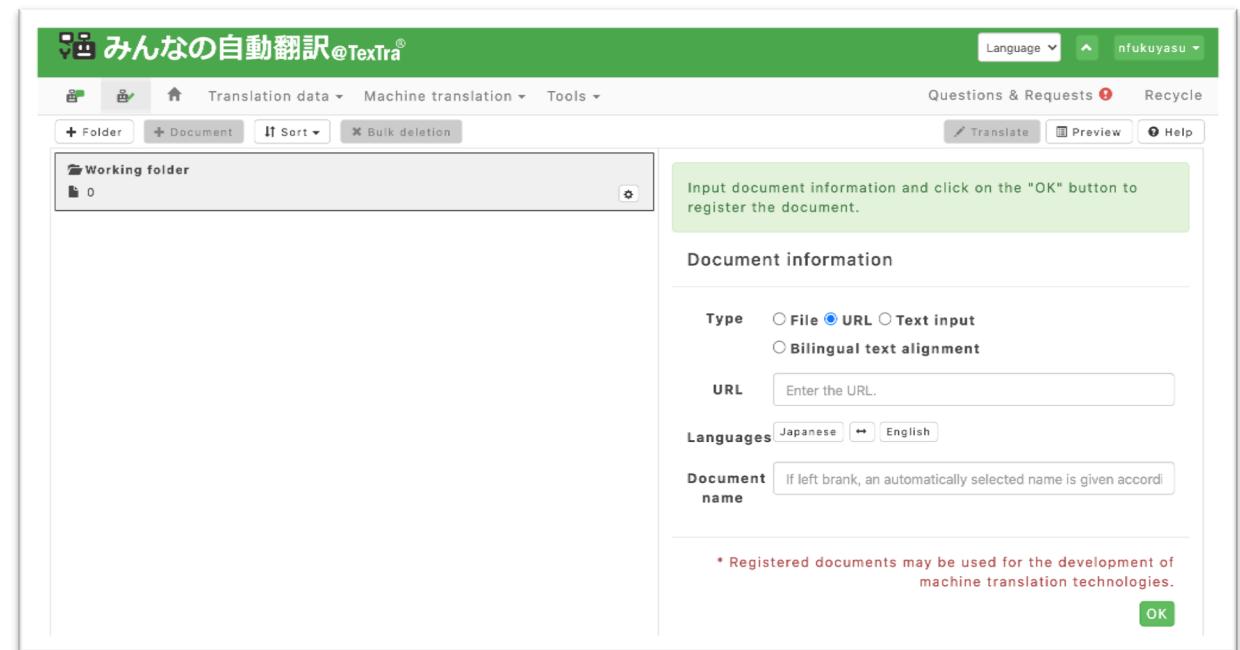
- Microsoft Word, Excel, Power
- XHTML and HTML
- Open Document Format (LibreOffice, OpenOffice.org)
- MediaWiki (Wikipedia)
- Plain text

Powerful tools

- Unicode (UTF-8) support: can be used with non-Latin alphabets
- Support for right-to-left languages
- Integral spelling checker
- Compatible with other translation memory applications (TMX, TTX, TXML, XLIFF, SDLXLIFF)

Available Tools: TexTra Translation Editor

- › [An online tool](#) provided by National Institution of Information and Communication Technology (NICT). Not an open source tool but anyone can use this for free.
- › Take advantage of machine translation engine
- › No need to install software on your PC (completes within the web browser)
- › Capable to translate many different file formats (just like OmegaT)
- › Capable to use WYSIWYG mode (not available with OmegaT)
- › Capable to create TM in multiple different file formats.
- › Capable to generate a glossary from the TM.



(My future dream) Translation Engine optimized for OSS usage

- Training the machine translation engine optimized for open source

- › By using the tools by using the community TM we should be able to get a lot better productivity.
- › With a machine translation engine optimized for open source usage, the productivity can become much higher. ([TexTra](#) is actually one good option for this)
- › Build an ecosystem around the machine translation (commercial/open source tools for CMS, Git repository, Instant messaging Social media etc.)

After all, “people” matters most: how can we make the translation work fun and productive?

- › Build a process to share the translation results and memory
 - › Mindset: translation as a fun collaborative work.
 - › Common Repository: not only for the translation outcome but TM as well.
 - › Work Process: it would be good to have a maintainer of translation projects
 - › Legal terms: CLA? DCO?
- › Community translation can become fun work?
 - › Learning through translating – translation is an opportunity to learn technologies.
 - › Collaboration with others (feedback to work, reputation, appreciation)
 - › Focus on the high value-added work and let the tool take care the low value part.

One final word:



Universal community translators, unite!



Universal community translators, unite!
(and share the translation memory)

THANKYOU!



Legal Notices

The Linux Foundation, The Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation's Trademark Usage Policy at <https://www.linuxfoundation.org/trademark-usage>, as may be modified from time to time.

Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute's trademark usage page at <https://lmi.linuxfoundation.org> for details regarding use of this trademark.

Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.

TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates.

Facebook and the "f" logo are trademarks of Facebook or its affiliates.

LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.

YouTube and the YouTube icon are trademarks of YouTube or its affiliates.

All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.

The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at <https://www.linuxfoundation.org/privacy> and its Antitrust Policy at <https://www.linuxfoundation.org/antitrust-policy>, each as may be modified from time to time. More information about The Linux Foundation's policies is available at <https://www.linuxfoundation.org>.

Please email legal@linuxfoundation.org with any questions about The Linux Foundation's policies or the notices set forth on this slide.