Collaborative Translation Efforts (CTE)
Trial Report

Feb, 2019

Vision of CTE
(Tentative, TBD)

More productive translation in "Open Source" way;
with more **Openness**,more **Collaboration**,
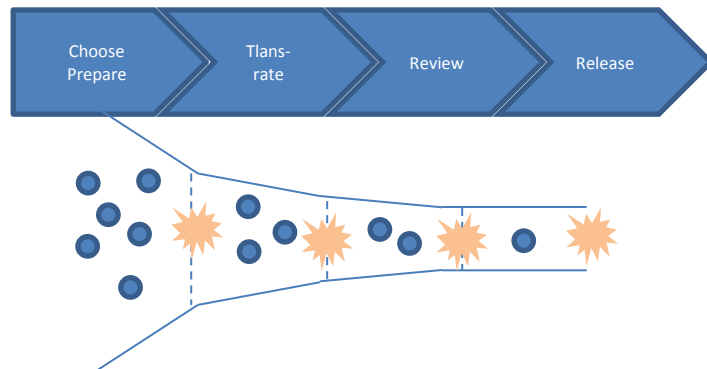and much more **Fun**!

Contents

1. Challenges in translation (especially in OSS resources)
2. Aim and approach
3. Trial
4. 1st Trial : By single person.
5. 2nd Trial : with 2 persons collaboration
6. Thought so far
7. Knowledge and feeling
8. Next steps

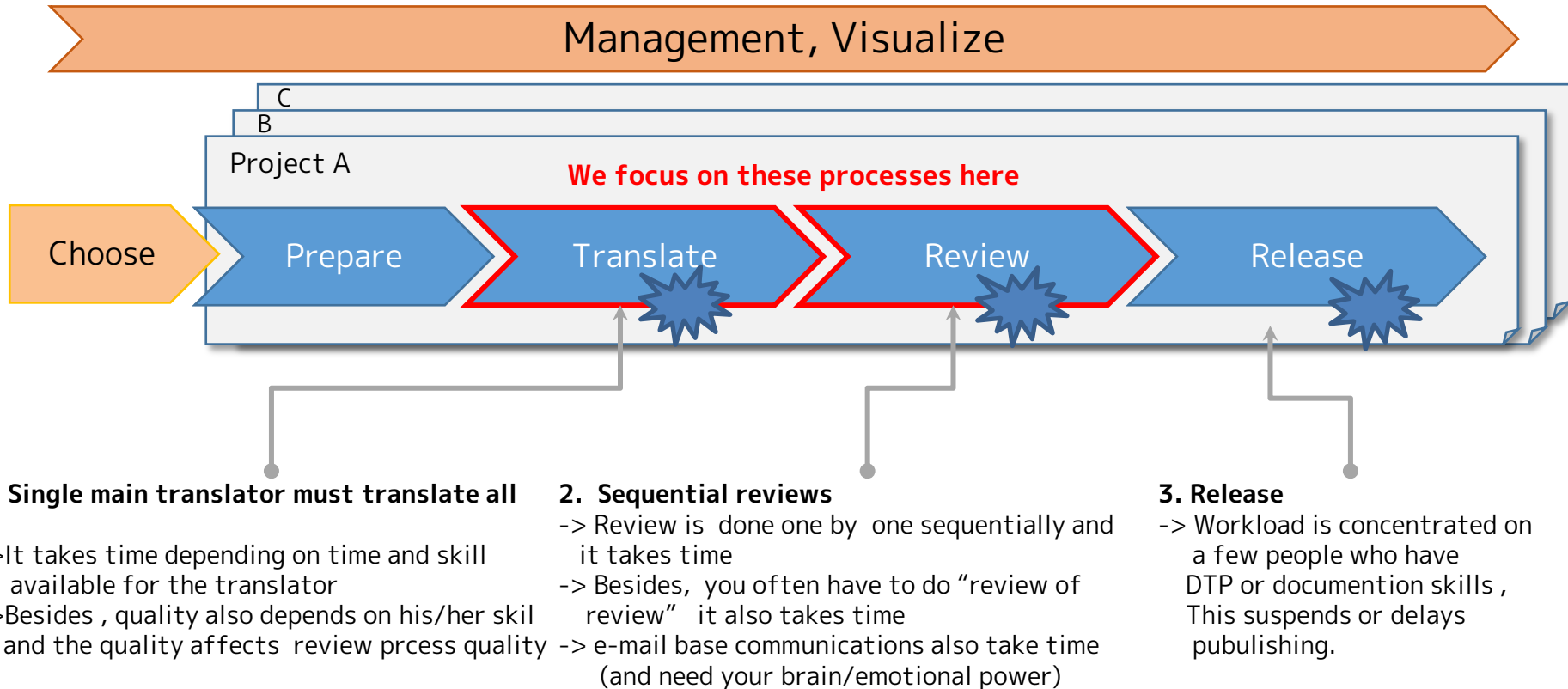# Challenges in translation (especially of OSS resources)

◆ **Question of "Quality**": Dilemma b/w OSS expertise and translation skills

- The OSS related translation strongly requires OSS expertise such as technology and its culture. It tends to be hard even for professional translators

- Meanwhile, if you have good enough undestanding about OSS, translation quality may not be good enough

◆ **Question of "Quantity"**: Unscalable prcess

- There exist bottlnecks for each process as in; Choose ,Prepare,Translate,Review,and Release
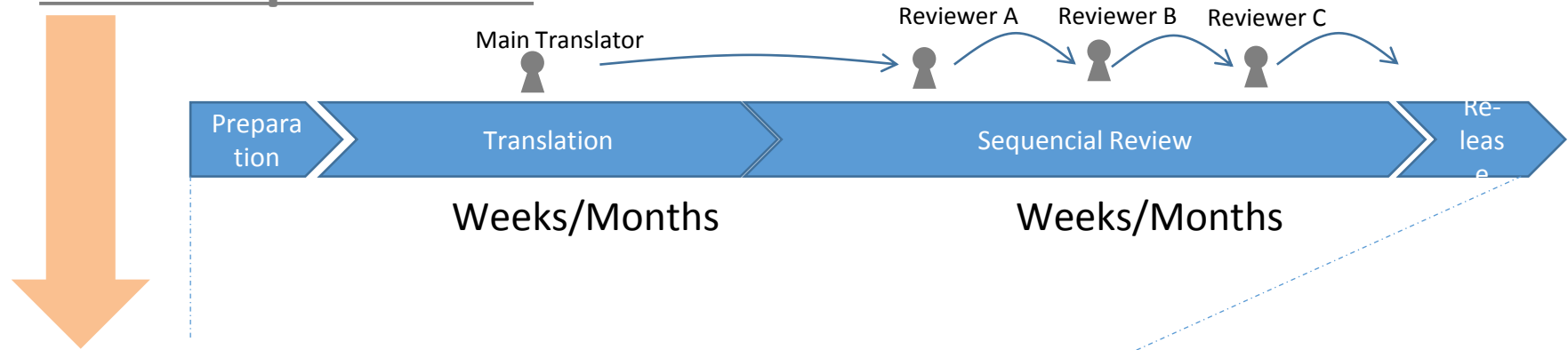
Major bottlenecks though traslation procecess

Management, Visualize

C

B

Project A

We focus on these processes here

Choose → Prepare → Translate → Review → Release

1. Single main translator must translate all

->It takes time depending on time and skill
    available for the translator
->Besides , quality also depends on his/her skil
    and the quality affects review prcess quality

2. Sequential reviews
-> Review is done one by one sequentially and
    it takes time
-> Besides, you often have to do "review of
    review"   it also takes time
-> e-mail base communications also take time
        (and need your brain/emotional power)

3. Release
-> Workload is concentrated on
    a few people who have
    DTP or documention skills ,
    This suspends or delays
    pubulishing.

Aim and approach

- **Aim:** **Focusing on "Question of Quantity"** (1 and 2 in the previous page), create virtuous cycle to "Question of Quality"
  - To be "prolific", above all
  - Invigorate peer review, make it more fun
  - The more collaborators gather, the more prolific (and scalable)
- **Approach:** Automation of Translation prcess and Review in realtime using tools
  - Tool 1: **Google Machine Translation**

    -> Shorten the time in Translation prcess, and Focus on Review process
  - Tool 2: **Hackmd**

    -> Simultaneous, realtime review (Edit markdown contents)
  - Toolrea 3: **Slack**

    -> Simultaneous, realtime review (Communication)
- **Evaluation indicators：**

1. Progress rate	2. Efficiency
Words /day	Minuites/(person*word)
Bigger, better	Smaller,better

Aim and approach (Cont'd)

Current processes

Main Translator

Reviewer A Reviewer B Reviewer C

| Preparation | Translation | Sequencial Review | Release |

Weeks/Months   Weeks/Months

CTE approach

Preparation
(None in some case)

Translation
（Machine）

Reviewer A, B, C

| | | Simultaneous Review (Realtime) | Release |

(GOTO Next Project)

Seconds/Minuites   Weeks
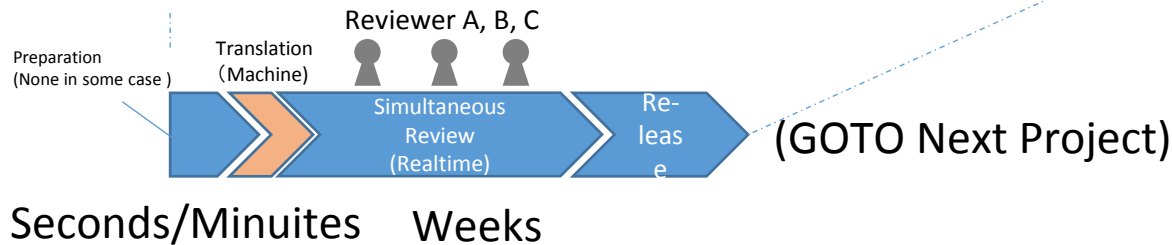
Trials: Basic policy
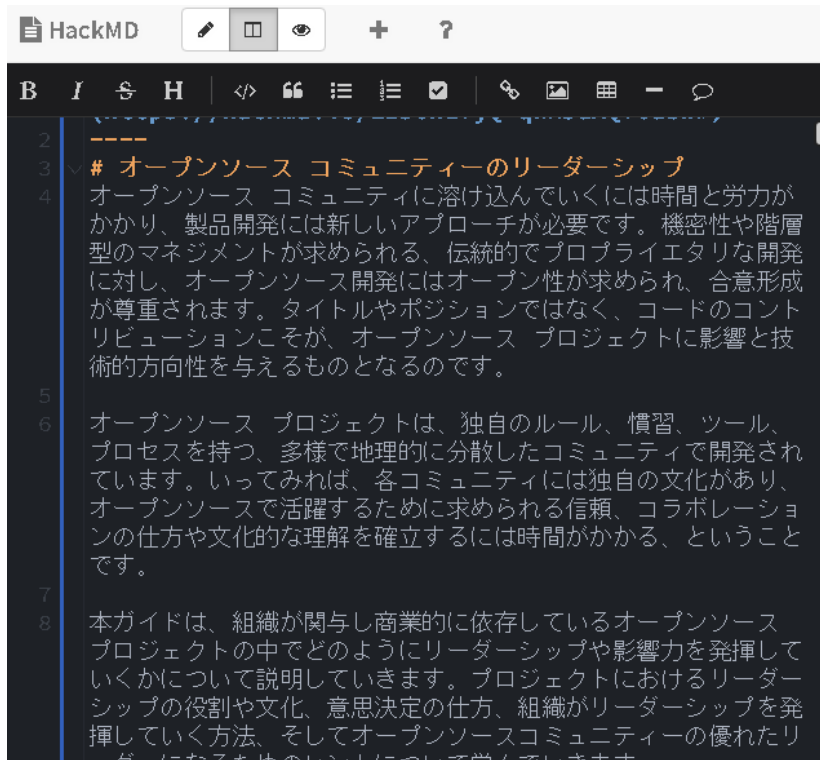
- **Measure and record time in your activity**
  -> This is an important factor for evaluation indicators
  -> NOT necessary to be so accurate

- **Online cross reviews should be done periodically , in short time**
  -> Eliminate the time you think hard on your own.
  -> Do not disucss for long time. Let's make reviews more casual
  -> Abstract and manage ToDo list

- **Online review should be done via Chat tool**
  -> Chat is more casual for many. We need more collaborators!

- **Above all, make it fun!**
  -> Enjoy original contents (Author's idea, thought)
  -> Enjoy interaction (Off topic is also important sometime)
  -> Enjoy progress (We are coming to the goal!)

1st Trial (by 1 person)

- Objective : Measure and evaluate how much single person takes
- Target Resource :
  - TODO Group "Building Leadership in an Open Source Community"
  - Approx.3600 words (24,000 letters)
- # of person : 1 (Taniguchi)
- Started: 28th Oct, 2018
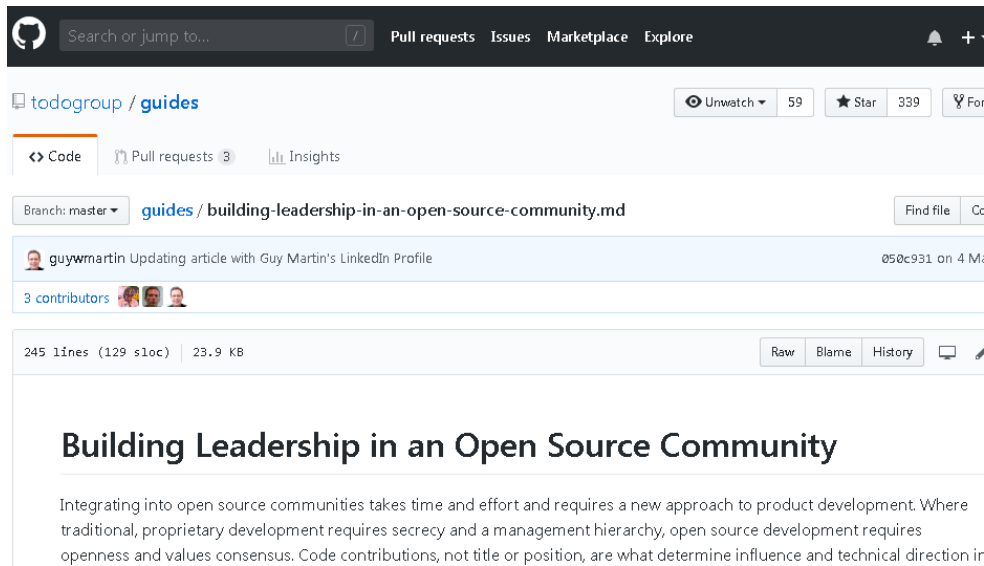
1ˢᵗ Trial (by 1 person):　Work in progress
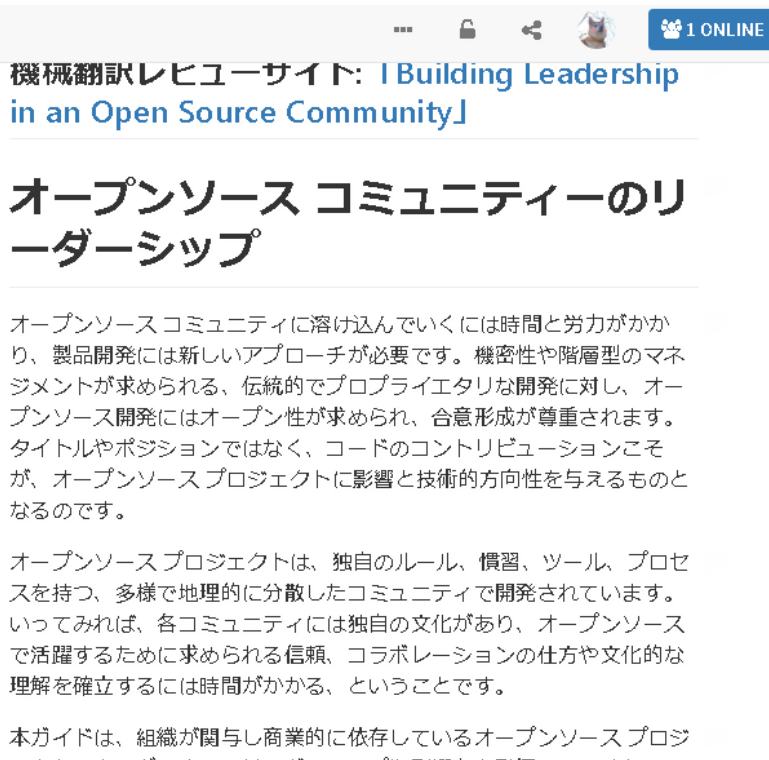
Edit pane(markdown)
*Paste machine translation output here at first

View pane

1st Trial (by 1 person): Result

- Outout:
  "オープンソース コミュニティのリーダーシップ"
     https://hackmd.io/aibsz3_JTqStRbyTdVO7rA
- Duration: 32days (28th Oct, 2018- 26th Nov, 2018)
  - Translation：Google Machine Translation   -> **30 min.** (manual work a little)
  - Review   -> **870 min.** (14.5h)
  - Review (correction by tool)   -> **30 min.**
  - Release (Image linking)   -> **30 min.**
  - **Sum**   **960 min.** (16h)
- Evaluation:
  - 1. Progress rate   ：**112.5 words/day**
     (3600 words/ 32days)
  - 2. Efficiency (Translation + Review) ：**0.26 min/(person*word)**
     (930min/1 person/3600 words＝0.26)

Note: Evalutation scores become worse because review must be done more than 1 person essentially

1ˢᵗ Trial (by 1 person): Result(Outcome)

https://hackmd.io/zgthoZZcTI-s3JXAg1pgkw?view

2ⁿᵈ Trial (2 persons)

- Obejective: Evaluate effectiveness by 2 collaborators
- Target resources:
  - LF「Certification Preparation Guide 」
    （1）DL site with introduction (104 words, 591letters)
    （2）PDF Slide (w/ 21 slides, 4,212 words, 27,799 lettes)
- # of person: 2 (Mieko-san and Taniguchi)

  ※Besides that, Inou-san from NEC Solution Innovator joined and did technical check as an engineer
- Started : 27ᵗʰ Dec 2018

DL site

PDF slide

2ⁿᵈ Trial (2 persons) Work in progres

Realtime multi-user editing on Hackmd   Chat on Slack

2nd Trial (2 person): Result

- Output:
  "Linux Foundation 認定試験ガイド"
  （1）DL site   https://hackmd.io/LCiwQGyfQKysQBVQDj61kw
  （2）PDF Slide   https://github.com/maabou512/CertPrepGuide
- Duratiom：
  （1）**Half a day** (27th Dec 2018 - 27th Dec 2018)
  - Translation: Google Machine Translation   -> **5 min.**
  - Review   -> **45 min.**
                                                   Sum   50 min.

  （2）**32 days** (28th Dec, 2018 – 28th Jan, 2019(Incl. Release process by InDesign DTP)
  - Translation: Google Machine Tranlsation   -> **30 min.** (mainly Copy and Paste)
  - Review（Self review)   -> **1140 min.** (19h：609min for Taniguchi), 540 min for Mieko-san)
  - Review（Online cross-review)   -> **550 min.** （5 times)
  - Release (DTP by InDesign)   -> **420 min.**
                                                   Sum   2140 min. (35h)

  - # of ToDo in online review: approx. 20 (closed all)

15

2nd Trial (2 persons) :Result(cont'd)

- Evaluation

  (1) DL site
  - Progress rate : **240 words / day**
    (104 words/0.5 day)
  - Efficiency (Translation + Review) : **0.24 min /(person*word)**
    (50 min/2 persons /104 words=0.24)

  (2) PDF Slide
  - Progress rate : **131 words / day**
    (4,212 words/32days*)
  - Efficiency (Translation + Review) : **0.20 min. / (person*word)**
    (1,720 min/2 persons/4,212 words=0.20*)

  * Note: Acutually , above scores includes "Release" activity , so scores must be (much) better

2ⁿᵈ Trial: Result (Outcome)

DL site introduction

PDF Slide (Prep Guide)

Trials Round-up

	1st Trial	2nd Trial (1)	2nd Trial (2)
# of Person	1人	2人	2人
Words	3,600	104	4,212
Translation time [min.] (Machine Translation	30	5	30
Review time [min.]	960	45	1690
Result 1. Progress rate [words/day](bigger better)	**113*1**	**240*2**	**131*2**
Result 2. Efficiency [min./(person*word)] (smaller better)	**0.26*1**	**0.24*2**	**0.20*2**
Special notes	-	-	Abstracted approx.20 ToDos (All closed)

*1 : Evalutation scores become worse because review must be done more than 1 person essentially

*2 : Acutually , above scores includes "Release" activity , so scores must be (much) better

Thoughts (so far)

- This approach (Machine Translation + Realtime cross review) has possibility about;
  - **Shorter leadtime** (Observed 1 -> 2 person)
  - **More efficiency** (ditto)
  - And **Quality improvement** (More issues and ToDos can be abstracted and tackled on accurately)

So, it can be exected that the more we collaborate, the better translation in terms of quickness, efficiency, and quality.

The diagram we intended in P7 (see right) is not comparable directly, but it can be effective indirectly.

Manual Translation Process v.s. Machine Translation Process (Taniguchi's idea)
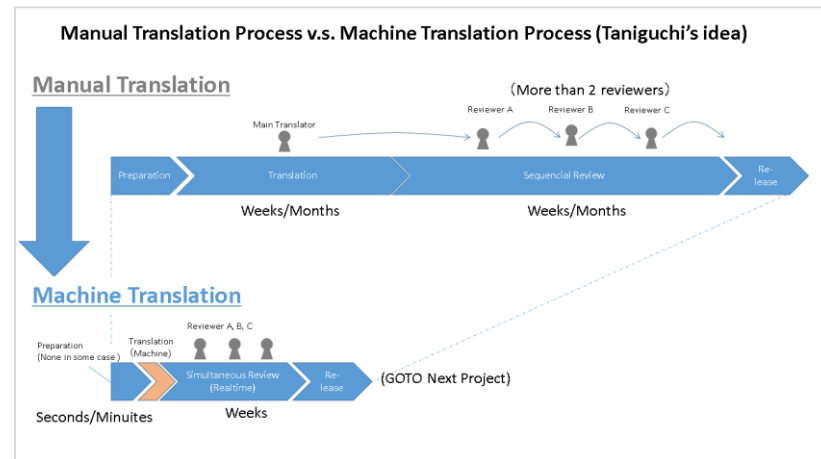
Knowledge and Comments

[About processes]
- **Ease of review.** We don't need to take "who translate this?" in consideration. Machines did it, you don't care of who did.
- Review process can be broken down like below (just idea basis)

Choose Prepare	Translate	Review	Release

Contents review (Self and cross review)

TODO management

Output definition (File format, etc)	Assign	Define Words	Cosmetic issues
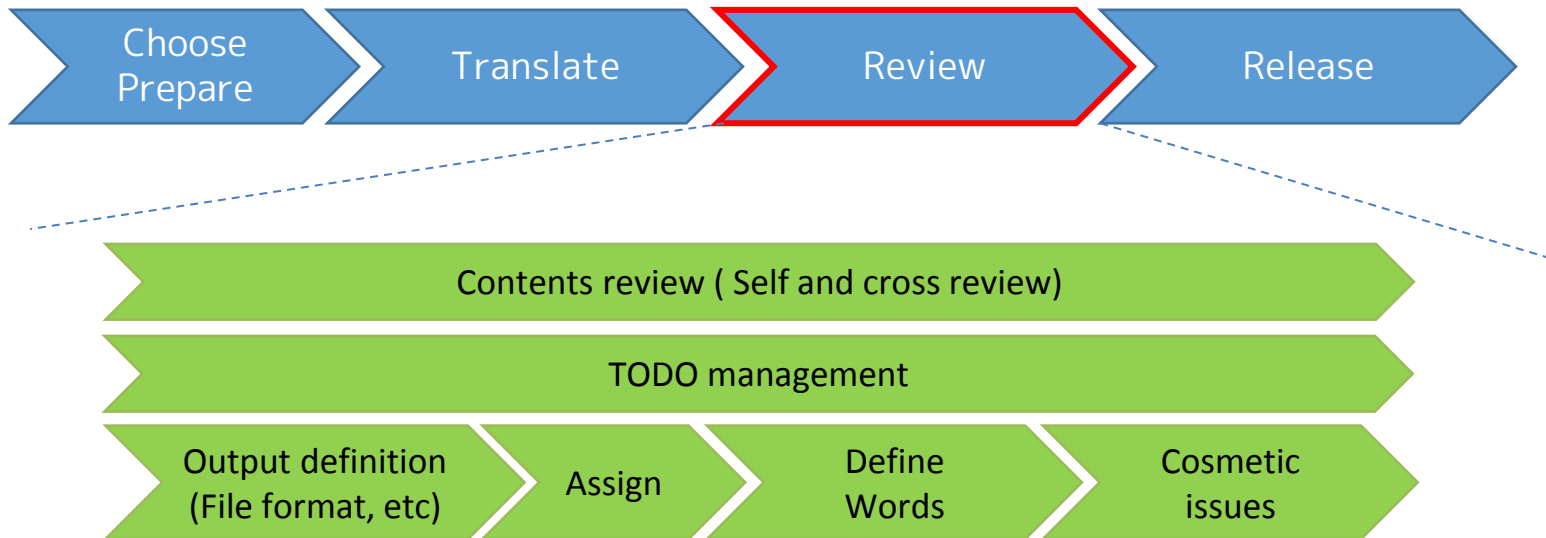
Knowledge and Comments

[Google machine translation]

- Very interesting tlanslation experience. (e.g . Same original sentences not always to be translated into same translation/It does mistake, but not like human)
- Basically , Quality is very low, **99%** of its output must be modified ultimately.
- **Characteristics of Machine Translation**
  1. Context is not understood semantically, so sometimes it can be even unreadable
  2. Subjects are mistaken or missing sometimes
  3. "De-Aru" and " Desu-Masu" style often mixd
  4. Very weak about "metaphor"
  5. Expressions are not coherent though a material
  6. Errors human beings never do

Next steps (Incl. Issues)

- **Improve processes** -> 3rd Trial ? What if 3 – 4 persons?
  - Shorten leadtime (Plan in advance and book members' schedules)
  - More efficient (Automate progress measurement, etc.)
  - Quality improvement (Create Check list, and so on)
  - And above all , much more fun!
- **Improve tools**
  - **Hackmd** -> Enhance evaluation indicator (Work time measurement) , Chat addon, Response improvement (Cloning Tokyo region), etc.
  - **Slack** etc. -> Function for ease of ToDo mgmt.
  - **Indesign** -> Realtime design check (not supported so far)
- **How to improve Machine Translation's quality?**
  - We "teach" good translation in OSS domain.
  - Can we have so-called "Reinforcement Learning" process?

Other Resources

- 「Collaborative Translation Effort（CTE）2nd Trial Spec (Japanese)」
  https://hackmd.io/xhHAQBZcTmmgEGRNgrzNhQ

  →Motivation and background about this trial and how to do it