



Crowdsourced translation practices

Alan Morera, Masaaki Collins, David Filip, Reinhard Schärer

Two Types of Patterns



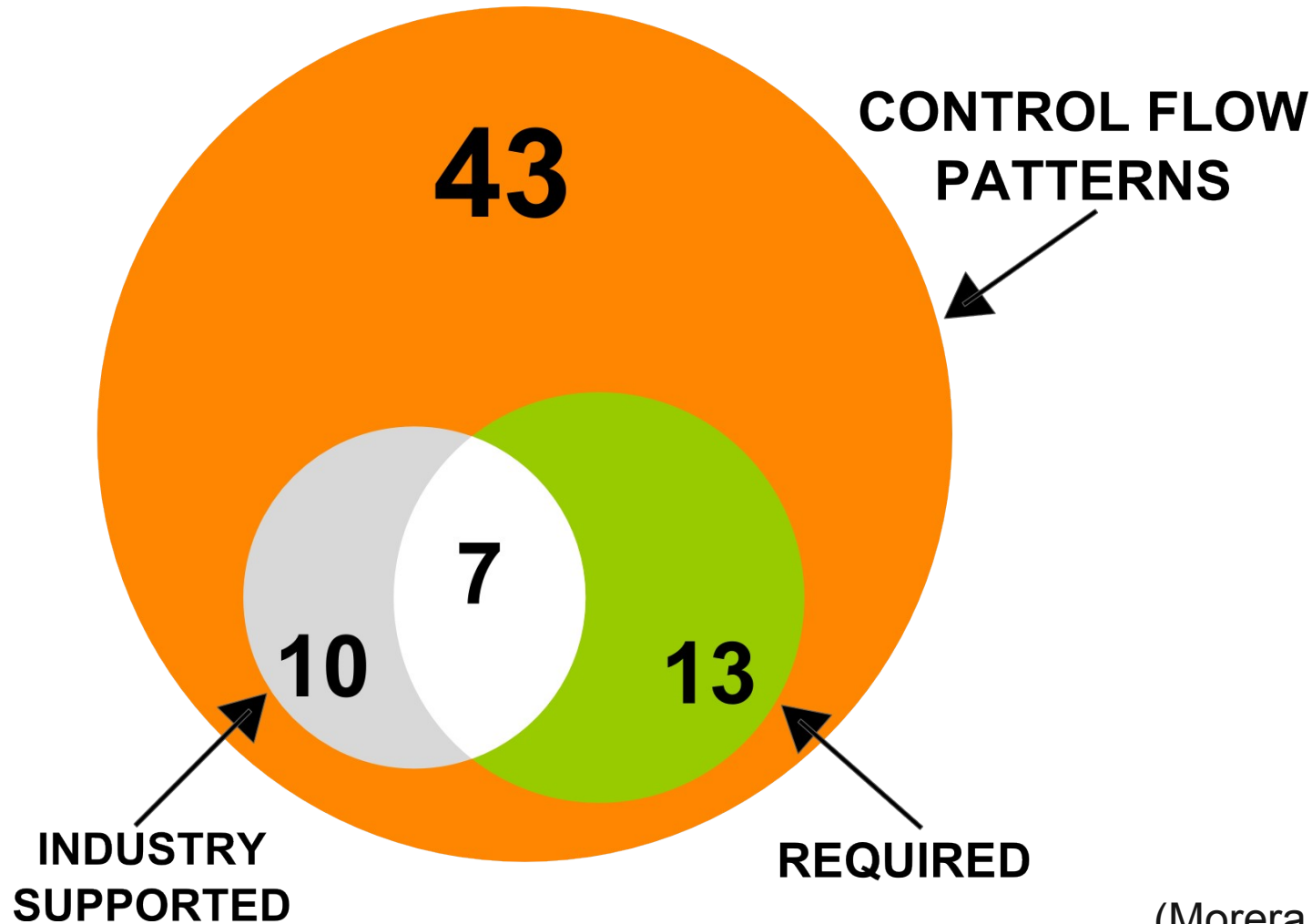
- Design Patterns: Solutions to recurring problems (Alexander et al, 1977)
- Workflow Patterns: A “series of constructs that are embodied in existing offers in response to actual modelling requirements” (Russel, 2005).

Workflow Patterns in Mainstream Tools vs. Crowdsourcing Tools



- Crowdin, AsiaOnline Wikipedia translation project, Facebook and Pootle
- GlobalSight and Idiom WorldServer

Workflow Patterns in Industry Tools vs Crowdsourcing Tools



(Morera et al, 2012)

Sequence



Sequence

- An activity in a workflow process is enabled after the completion of a preceding activity in the same process(Russell *et al*, 2006).

Sequence workflow

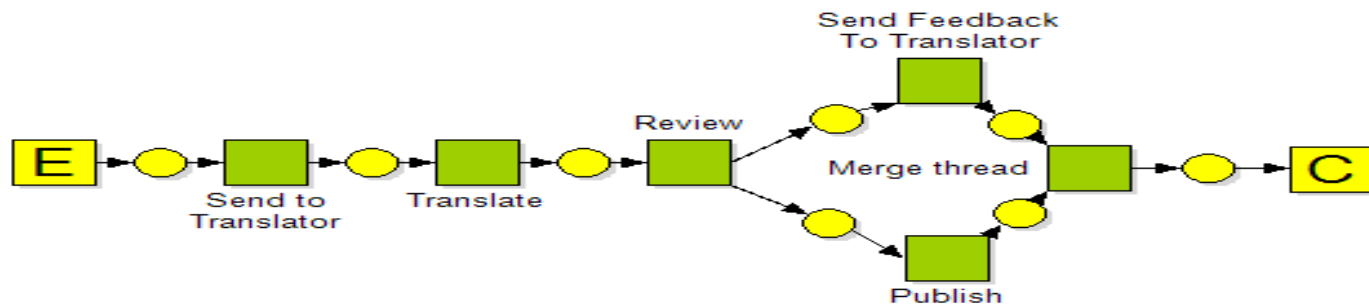


Parallel Split



- The divergence of a branch into two or more parallel branches each of which execute concurrently.(Russell *et al*, 2006).

Parallel Split Workflow

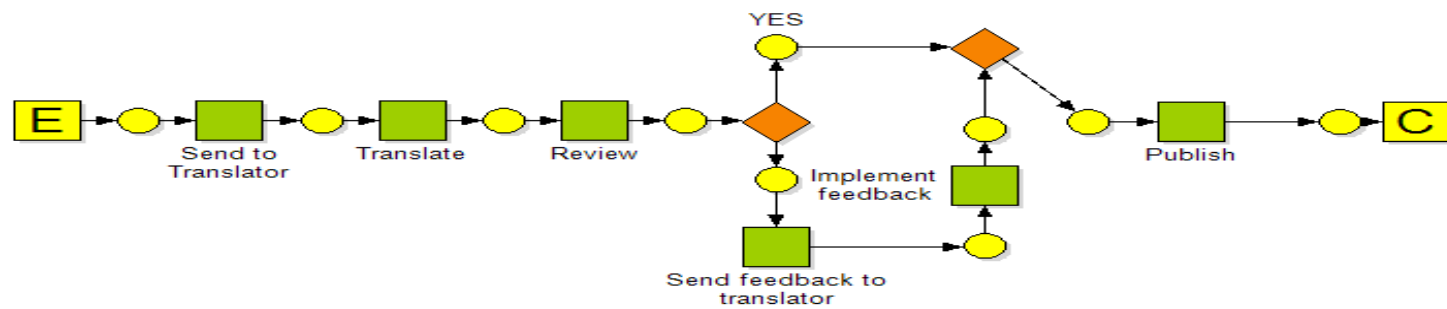


Exclusive Choice



- The divergence of a branch into two or more branches. When the incoming branch is enabled, the thread of control is immediately passed to precisely one of the outgoing branches based on the outcome of a logical expression associated with the branch (Russell *et al*, 2006).

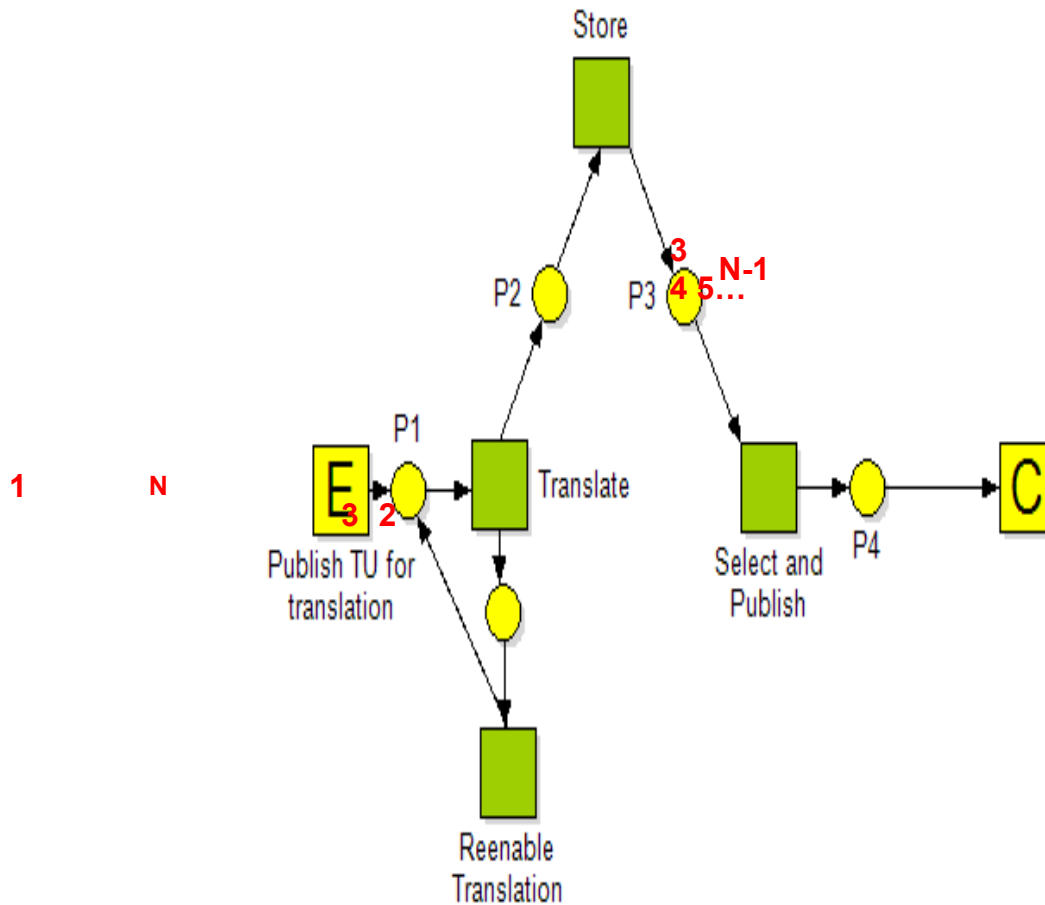
Exclusive choice



Multiple Instances Without Runtime Knowledge Pattern

- Within a given process instance, multiple instances of an activity can be created. The required number of instances may depend on a number of runtime factors, including state data, resource availability and inter-process communications and is not known until the final instance has completed (Russell *et al*, 2006).

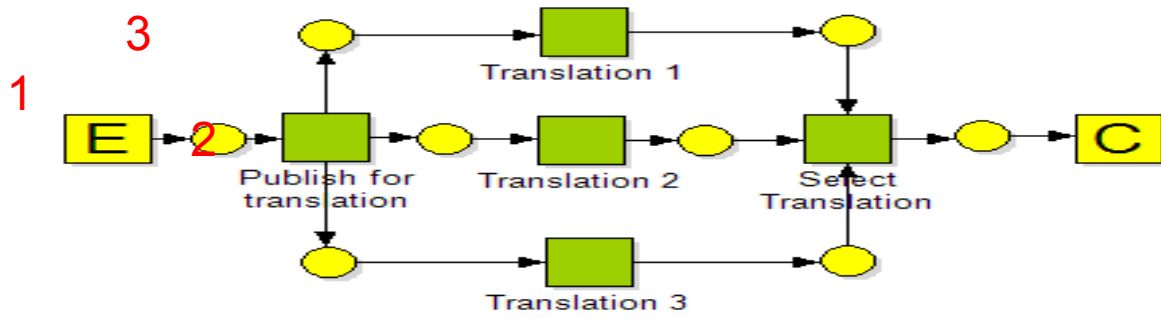
Multiple Instances Without Runtime Knowledge Pattern



Multiple Instances With Design Time Knowledge Pattern

- Within a given process instance, multiple instances of an activity can be created. The required number of instances is known at design time. These instances are independent of each other and run concurrently (Russell *et al*, 2006).

Multiple Instances With Design Time Knowledge Pattern





The practices

Click to edit Master subtitle style

16 Practices



- Kiva, Facebook, DotSub, Crowdin, Pootle, Asia Online Wikipedia project, Amara and Launchpad's Rosetta.
- Looked at translation payload and metadata generating processes.
- No resourcing, motivation or control practices
- 14 Practices identified, added two from collaborative-translation-patterns.com
- 8 Interviews between 40 and 87 minutes.
- Over 64000 words.
- NVIVO 10 Coded

Sixteen Practices



Open Alternatives

The creation of redundant alternative translations for a single TU which are visible to the contributors and often not limited in number.

Open Alternatives



- Enables Open redundant assessment.
 - Having alternatives enables *“the voting mechanism [...] that's of course an advantage that will lead to higher quality”*.
- Fosters crowd engagement.
 - *“the advantage of letting contributors make more than one suggestion is that there's this community dynamic”* because *“as soon as you have more than one translation for any given string, people can start discussing 'why is this the better translation, why that...’”*
- Potentially highly redundant (wasteful)
 - *“people translate the same stuff over and over again”*

Hidden Alternatives

The creation of redundant alternative translations for a single TU which are not visible to the contributors and are often limited in number.

Click to edit Master subtitle style

Hidden Alternatives

- Enables frequency-of-translation based selection.
 - *“when you see [a translation] repeated exactly the same way... Frequency of occurrence is a check mark that 'Yes, go ahead. No human approval needed”*.
- By hiding the translations community activities such as voting and commenting on them disappear and this weakens the motivation related to those activities and must be compensated for with rewards.
 - *“If it's a one way street, where users only submit their stuff and do not see what others have submitted, only the organizer sees this, I'm wondering what happens to my motivation”*
 - *“their contribution earned them some points, earned them lottery tickets, whatever it was and that's all they cared about”*
- Improves resource-to-task ratio (less wasteful)
 - *“It's kind of a waste if you have people translate the same stuff over and over again”*



Open

Assessment

The collection of redundant assessments, often in the form of votes that are visible to the

contributors.

Click to edit Master subtitle style

Open Redundant Assessment



- Open assessment enables vote based selection, which results in a more crowd-managed process.
 - *“The advantage of it is that you don't have to have a dedicated person within your organization to choose the correct translation”*
- This type of assessment opens another venue for crowd engagement, which is valuable for marketing purposes.
 - *“you can create a much more lively community interaction”*



Metadata Based Selection

The automatic selection of a translation
from a collection of alternatives with the
guidance of metadata.

Click to edit Master subtitle style

Metadata Based Selection



- Automated selection can be the only alternative in extremely big projects.
 - *“if you have 100 page document that you are translating through crowd, maybe computer selection is not the best way to do it. If you've got a million page document, maybe the computer selection is the only way that you are gonna do it.”*
- Automated selection can be tricked.
 - With automated selection you run the risk *“things that you do not want to see in your translation may be in there”*.

Metadata Based Selection

- Four types of metadata based selection:
 - Frequency of translations if you have used hidden redundant alternatives.
 - You could approve suggestions *“when you see it repeated exactly the same way. Frequency of occurrence is a check mark that 'Yes, go ahead. No human approval needed’”*
- Trusted translators if you are logging the performance of your translators.
 - You could approve suggestions *“Maybe based on the fact that the users is known or has constantly submitted good translations”*
- Crowd assessment both in the form of open or hidden assessment.
 - You could *“rely on a certain number of up-votes, of positive votes for a translation, so, if a translation gets 5+ votes it gets approved automatically”*
- Time stamp, under the assumption that the latest translation will be the best available.
 - When not translation has been upvoted *“the last added [translation] will be used”*

Metadata Based Selection

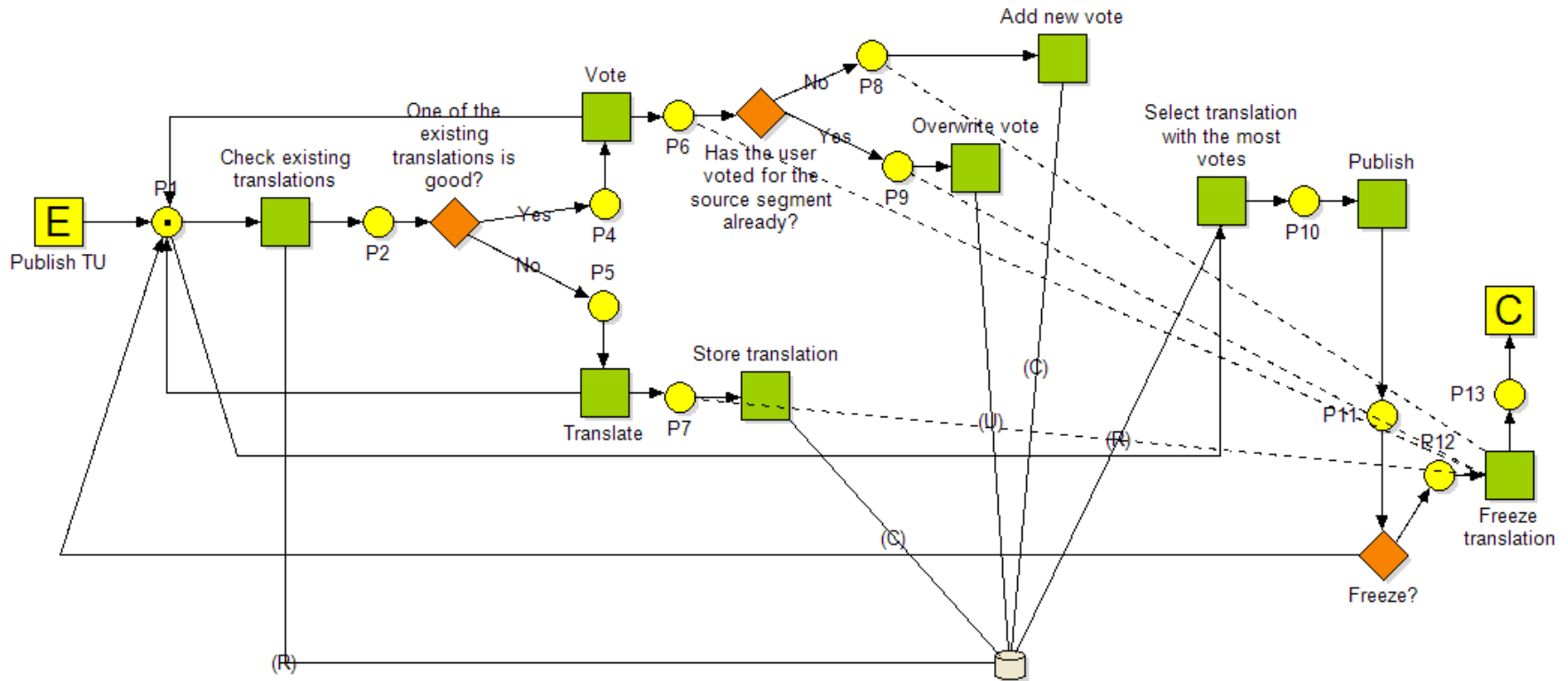


- Frequency of translations based selection if you have used hidden redundant alternatives.
 - *“maybe it's two or three... I don't know exactly what the algorithms were, but you know, when you see it repeated exactly the same way. Frequency of occurrence is a check mark that 'Yes, go ahead. No human approval needed'.*
 - Offers quality without the caveats of social influence bias (*Muchnik et al, 2013*)
- Open Assessment Based selection is optimal when the criteria for quality is acceptability
 - *“it's really clear which [translation] is better by considering the number of votes”.*
 - *“best possible translation as it is viewed by the people who do the translation”*
 - It is submitted to the caveats of social influence bias (*ibid.*)

Takeaways

- Patterns to enable higher engagement:
 - If as a company you are using crowdsourced translation in order to foster community engagement, you want to use open alternatives, open assessment and vote based selection because they enable the biggest number of people to be involved and give the crowd the most power. This can result in a slower process, but helps creating a community.

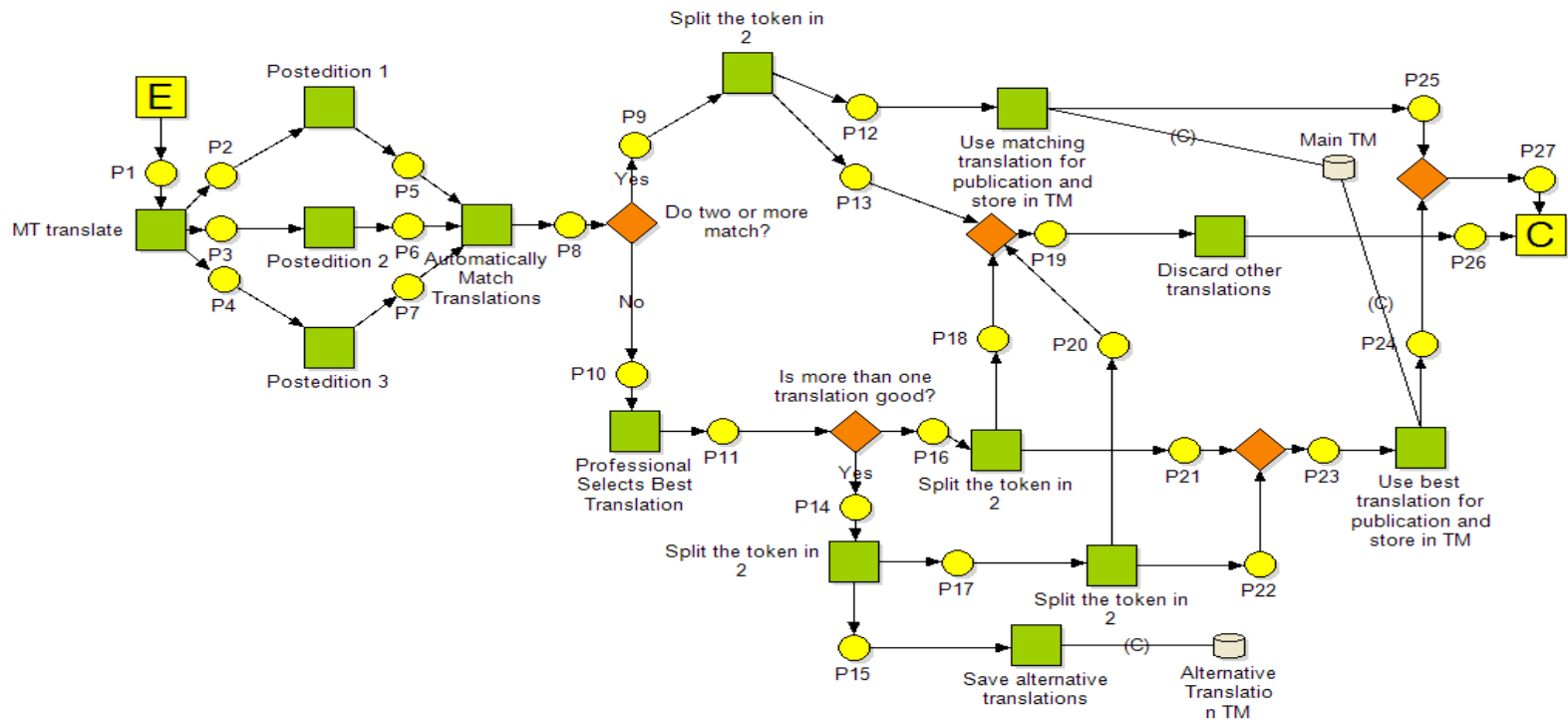
Facebook Community Process (2009)



Takeaways

- Patterns for increased throughput
 - If as a company you are using crowdsourced translation in order to translate a high volume in a short time, you want to use MT and TM combined with hidden redundancy and frequency based selection. This accelerates the process, but does not help creating a community.

Asia Online Wikipedia Project





Thanks!

Click to edit Master subtitle style