

**WIPO**  
WORLD  
INTELLECTUAL PROPERTY  
ORGANIZATION

**Pathways to integrating machine translation  
into translation workflows**

InnoVent, UNOG, Geneva  
January 16, 2020

Laurent Gottardo  
Head, Translation and Terminology Technology Section  
PCT Translation Division

WIPO FOR OFFICIAL USE ONLY

2

All roads lead to Rome...  
... but don't blindly trust your GPS.

**WIPO**  
WORLD  
INTELLECTUAL PROPERTY  
ORGANIZATION

WIPO FOR OFFICIAL USE ONLY

3

## What this talk is *not* going to be about

- Specific MT engines and the quality of their outputs with respect to language combinations, document types, etc.
- How to choose, build, or train MT engines.

4

## What this talk *is* going to be about

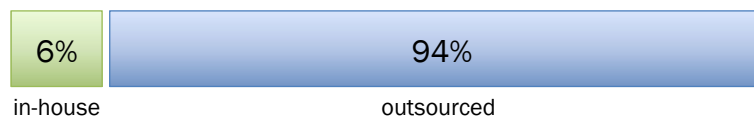
- **Why.** Integrate MT to reduce translation effort & costs, increase turnaround time, while maintaining or increasing translation quality.
- **How.** Integrate MT by streaming and distributing documents and their MT outputs in adaptive translation workflows to translators, revisers, and technical specialists where it makes sense from a business perspective and individual human skillsets.
- Birds-eye view of our MT translation workflow with hopefully lots of questions!

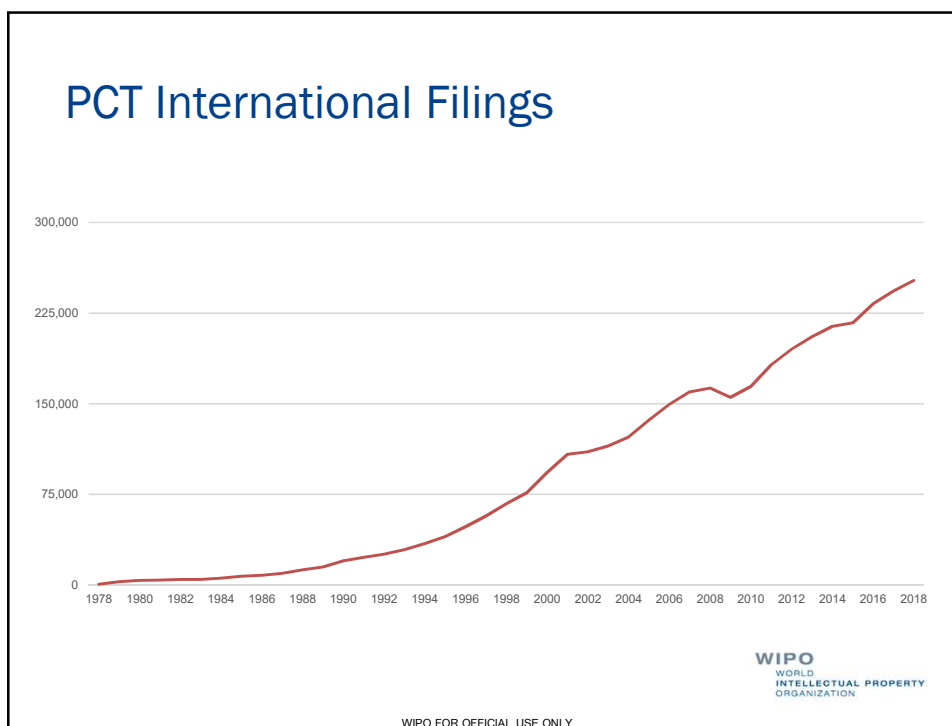
## PCT Translation Division

- The PCT Translation Division has over 30 years of experience in translating technical abstracts and patentability reports for publication.
- Applications may be filed for publication in 10 languages: Arabic, Chinese, English, French, German, Japanese, Korean, Portuguese, Russian, Spanish.
- The PCT Translation Division may handle up to 3 languages for each international application: the initial publication language and up to two further publication languages (English, French).

## Translation volume (2018)

- Approx. 167 million words  
(patent abstracts & patentability reports)
- Approx. 611,741 translations
- Steady increase in yearly translation volume





8

## Ingredients for MT workflow integration (1)

1. **Customizable translation workflows** – they should neither be carved in stone nor should it be expensive and time-consuming to change them, while the expertise to do so should reside in-house.
2. A **business-driven, rule-based streaming and quota-based distribution engine** to create translation streams.

WIPO  
WORLD  
INTELLECTUAL PROPERTY  
ORGANIZATION

WIPO FOR OFFICIAL USE ONLY

## Ingredients for MT workflow integration (2)

3. **Human element – Linguists and technical specialists** should be assigned to those workflow steps where it makes most sense given the nature of the data and their individual expertise.
4. If available, a machine translation **quality estimation score**. To help get a better sense of which documents could be effectively post-edited.

## 1. Translation workflows (1)

- Need to be able to **adapt to evolving business needs**. Workflows should be customizable, easily integrate with new functionalities, and be embedded into enterprise-level, scalable **translation management systems**.
- Should be **driven by business needs**. For instance, MT should *not* be applied at (vendor-)predefined workflow steps but rather where it makes sense from the business perspective, using linguists and technical specialists optimally.

## 1. Translation workflows (2)

- Specifically, you map translation processes using
  - Automatic workflow steps that perform predefined actions without human intervention.
  - Human workflow steps which necessitate human intervention.
  - Conditional branching between the two; and subworkflows.

## 1. Translation workflows (3)

- Example: Automatically and randomly select a percentage or fixed amount of documents to be quality-controlled after translation has been performed by a given supplier.
- Send those documents into a QC pool where revisers pick work in a self-service fashion.
- Automatically generate QC reports and share them with the suppliers.

## 2. Streaming and Distribution engine (1)

- A business-driven, **rule-based streaming engine** to create different translation streams of documents matching operational or document-related criteria.
- A **quota-based distribution engine** to automate common project management tasks and distribute work in a (semi-)automated manner matching translation and quality-control quotas per individual external translation suppliers.

## 2. Streaming and Distribution engine (2)

1. Rule-based streaming by
  - Document type, linguistic similarity, lang. comb.
  - Word count, TM pretranslation rate
  - Machine translation quality estimation score (MTQE)
  - Meta-data
2. Distribution to external translation suppliers based on
  - Translation quota
  - Quality-control quota

### 3. Human element

- Linguists & technical specialists have different skillsets / areas of expertise.
- They should be used in the most effective way, e.g.
  - technical specialists in a given domain
  - in-house translators on urgent requests
  - etc.

### 4. Machine Translation Quality Estimation

- Estimate the quality of an MT output by assigning a score (both on a sentence and document level) and use it in translation workflows to stream the translations into different pools.
- Classic supervised machine learning challenge.
- Trained on data triplets (collected since 2015)
  - Source, MT suggestion, human translation



17

## 4. Machine Translation Quality Estimation

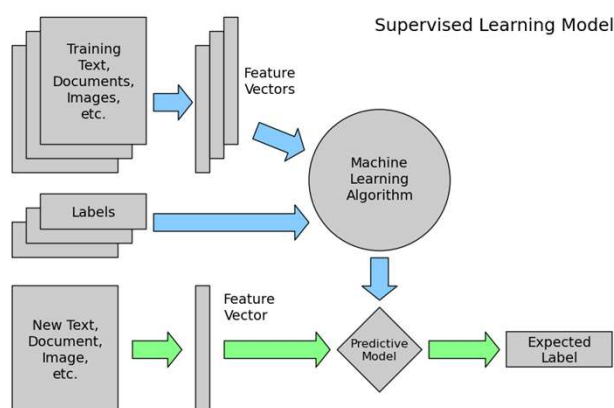
- Features were extracted from sentences
- Predictive model, classifier
  - Initially: Binary (good / bad)
  - Then: More nuanced algorithm providing a score between 0 and 1
- 3 rounds of human evaluation to determine the threshold up to which the score was accurate

WIPO  
WORLD  
INTELLECTUAL PROPERTY  
ORGANIZATION

WIPO FOR OFFICIAL USE ONLY

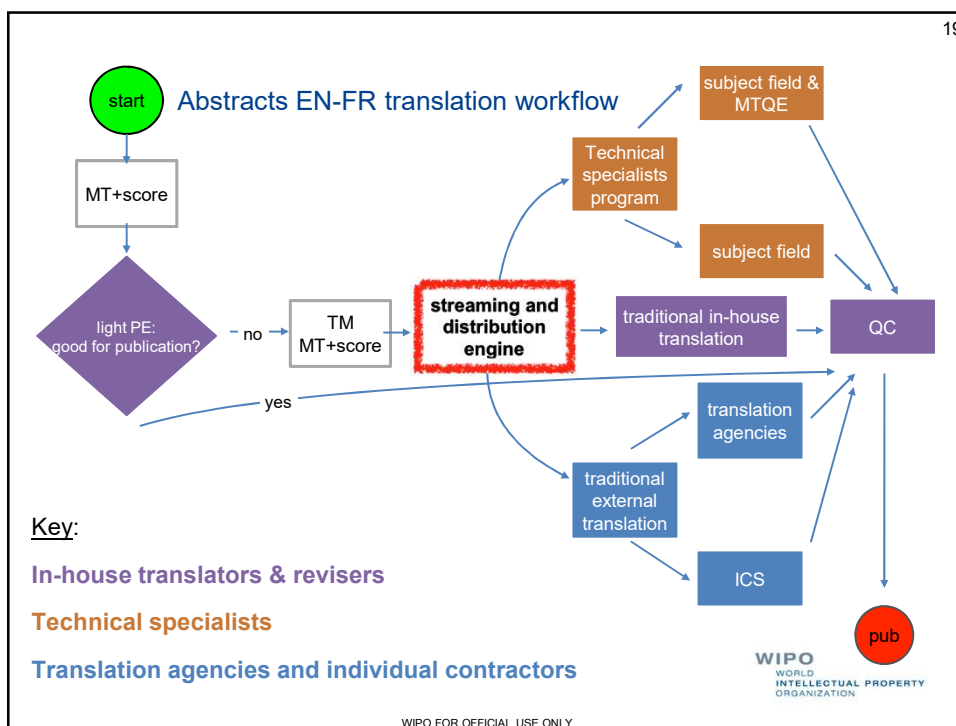
18

## 4. MTQE - Supervised Machine Learning



WIPO  
WORLD  
INTELLECTUAL PROPERTY  
ORGANIZATION

WIPO FOR OFFICIAL USE ONLY



20

## Benefits

Workflow step	Volume % of overall project	Productivity increase	Costs
Preselection (light post-editing) good for publication	5 to 10%	280%	n/a
Technical specialists program	15 to 20 %	n/a	37%
Traditional in-house translation	1.5 to 5 %	150%	n/a
External translation suppliers	70%	n/a	100%

WIPO  
WORLD INTELLECTUAL PROPERTY ORGANIZATION

WIPO FOR OFFICIAL USE ONLY

## Conclusion

- Business strategies should not be impacted or driven by (translation) software limitations.
- Machine translation is only one aspect of translation workflows.
- Translation workflows should be customizable.
- Linguists and technical specialists and their individual skillsets should not be randomly assigned documents for translation/post-edition.

**THANK YOU**

**QUESTIONS?**