

Training, feedback and productivity measurement with NMT and Adaptive MT

Smartcat

AMTA 2018

Why?

- Why MT — does it make sense for everybody?
- And if so, why do companies choose to train MT engines over existing generic ones?
- Why do we need to measure productivity and collect other data and does it impact our processes?

Smartcat

AMTA

Industry Challenges

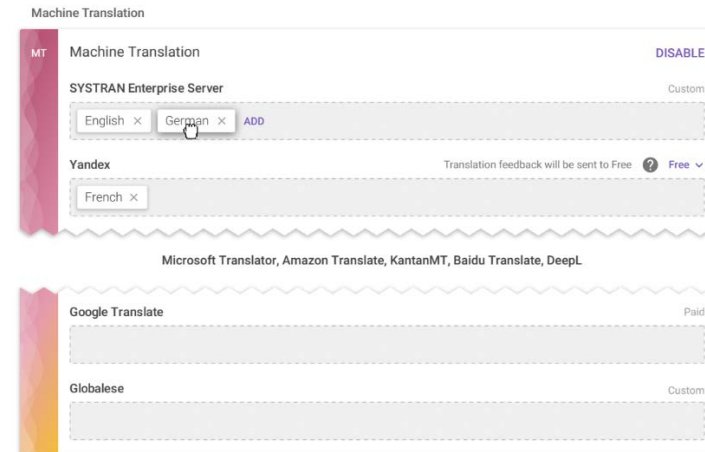
- **Many MT providers**
- **Lack of integrated benchmarks**
- **Little usable feedback to project participants**
- **Small organizations (LSPs and End-Users) left behind**
 - Generally no in-house expertise
 - Training engines takes time and data, and is expensive!

Smartcat

AMTA

Many MT Providers – Decisions, decisions...

- Various technologies each with their own advantages and flaws
 - RBMT
 - SMT
 - NMT
 - Adaptive MT
- Some MT providers add the option to train or customize engines
- CAT Tools do a low level integration of many of these MT engines or don't support them at all



Smartcat

AMTA

How do we choose?

- Most process participants simply default to the easiest option in their CAT tools
- Other rely on the opinion or preferences of their post-editors – but how reliable is that when for example NMT's fluency can distort an evaluation?
- Pricing is important!

Smartcat

AMTA

Lack of Integrated Scoring

- **Multiple MT scoring systems**
 - BLEU
 - METEOR
 - LEPOR
- **These scores are not available in CAT tools.**
- **Scores for engine trained on an unknown corpus can be unreliable and scores for generic engines vary with time.**

Smartcat

AMTA

Or Benchmarks!

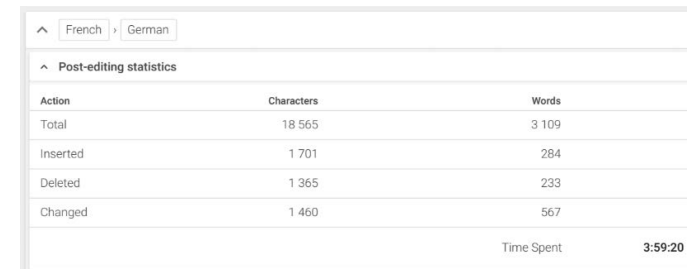
- **Quality evaluation approaches used for human translation cannot be used for quality evaluation of the machine translation.**
 - Do we create new rating systems (MQM comes to mind)
 - How do we make them available to participants?
- **What other metrics could assist project participants?**
 - Effort/Productivity/Prices
 - QA reports

Smartcat

AMTA

Solutions

- **Integration of benchmark in the CAT tool interface**
 - Display standard scores
 - And possible custom scores
 - Add pricing calculations to the mix!
- **Calculate and display PEMT productivity numbers:**
 - For each MT engine integrated in the CAT tool
 - For individual translators on various MT engines
- **Produce after-action reports**



The screenshot shows a software interface for post-editing statistics. At the top, there are language selection buttons for 'French' and 'German'. Below this is a section titled 'Post-editing statistics' which contains a table with three columns: 'Action', 'Characters', and 'Words'. The table lists four categories: 'Total', 'Inserted', 'Deleted', and 'Changed', with their respective character and word counts. At the bottom right of the table, there is a 'Time Spent' field showing '3:59:20'.

Action	Characters	Words
Total	18 565	3 109
Inserted	1 701	284
Deleted	1 365	233
Changed	1 460	567

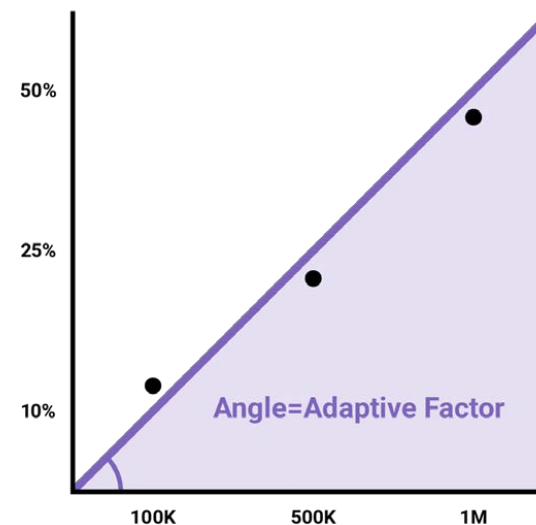
Time Spent 3:59:20

Smartcat

AMTA

The Adaptive MT Conundrum

- Standard scoring system cannot score the adaptiveness of the engine.
- Productivity metrics can still be produced but should vary more than with other engine types.
- Can we devise a new metrics
 - Adaptive Factor?
 - Other methods



Smartcat

AMTA

Feedback – Project Managers

How useful could the access to these metrics be for project managers?

- Select the best potential MT engine for each language pair/topic at the best price
- Use productivity and QA reports to assist in task assignment
- More accurate costs analysis around MT and PEMT
- Better deadline predictions

Smartcat

AMTA

Feedback – Translators

How about for the translators?

- Lower reluctance to use MT when presented with updated productivity numbers
- A new evaluation of the per word pricing
- In the context of a marketplace, more information for the potential buyers or higher ranking for the translator
- Improve training using reviewed “after action” reports and productivity benchmarks

Source text	Translation	Editing	Proofreading
In the town where I was born lived a man who sailed to sea.	In der Stadt, in der ich geboren wurde, lebte ein Mann, der zur See gefahren war.	In der Stadt, in-der wo ich geboren wurde, lebte ein Mann, der zur in See gefahren gestochen war.	In der-Stadt, wo-ich-geboren-wurde, meiner Heimatstadt lebte ein Mann Seemann, der in-See gestochen ausgesegelt war.

Smartcat

AMTA

Next Steps

Additional process automation – not only can project be created automatically but:

- The best possible MT resources can be allocated automatically by the system.
- Translators can be assigned to the tasks without human interaction to match deadline and cost requirements.
- Real-time productivity numbers can be used to re-allocate translation resources as needed

Smartcat

AMTA

Small Organization Access

- **Generally no in-house experience**
- **Training performant engines takes data, time and expenses**
- **Not practical for temporary needs**

Smartcat

AMTA

Solutions

Utilize the marketplace model:

- **Trained MT engine marketplace**
 - Real-time scores and productivity benchmarks
 - Data source information
- **Training data marketplace – already in place**
- **PEMT vendors marketplace – partially in place**
 - Real-time productivity benchmarks
 - QA reports and feedback

Smartcat

AMTA

Thank you!

Jean-Luc Saillard
Jean-luc@smartcat.ai
smartcat.ai

Smartcat

AMTA