

APPLYING AI IN TRANSLATION

HELPING GLOBAL COMPANIES TRANSLATE EFFICIENTLY

Uber

zendesk®

Whirlpool
CANADA

 MEMSOURCE

ABOUT MEMSOURCE

- Cloud-based translation management system
- Translation editors: browser, desktop and mobile
- Global customer base
- Founded in 2010
- Currently about 100 people (around 50 developers)
- AI & big data teams established in 2017

OUTLINE

- **Motivation: Enabling MT**
- Identification of Non-Translatables
- MT Quality Estimation
- Next Steps
- Conclusion

MOTIVATION: ENABLING MACHINE TRANSLATION

- MT quality has increased dramatically in the recent years
- MT adoption in translation industry is still limited
- MT is still not tightly integrated in translation workflow
 - Its use can be complicated

MAKING MT IN TRANSLATION SEAMLESS

- Pain points:
 - Building and maintaining of MT engines
 - Selection of MT engine for given content
 - Analysis of MT engine performance before translation
 - Updating MT engines with new data
- Goal:
 - Make MT even simpler to use than translation memories
 - Only then can the benefits be fully utilized

OUTLINE

- Motivation: Enabling MT
- **Identification of Non-Translatables**
- MT Quality Estimation
- Next Steps
- Conclusion

IDENTIFICATION OF NON-TRANSLATABLES

Web Editor Edit Tools Format Document Help

B I U X₂ X² Split Join

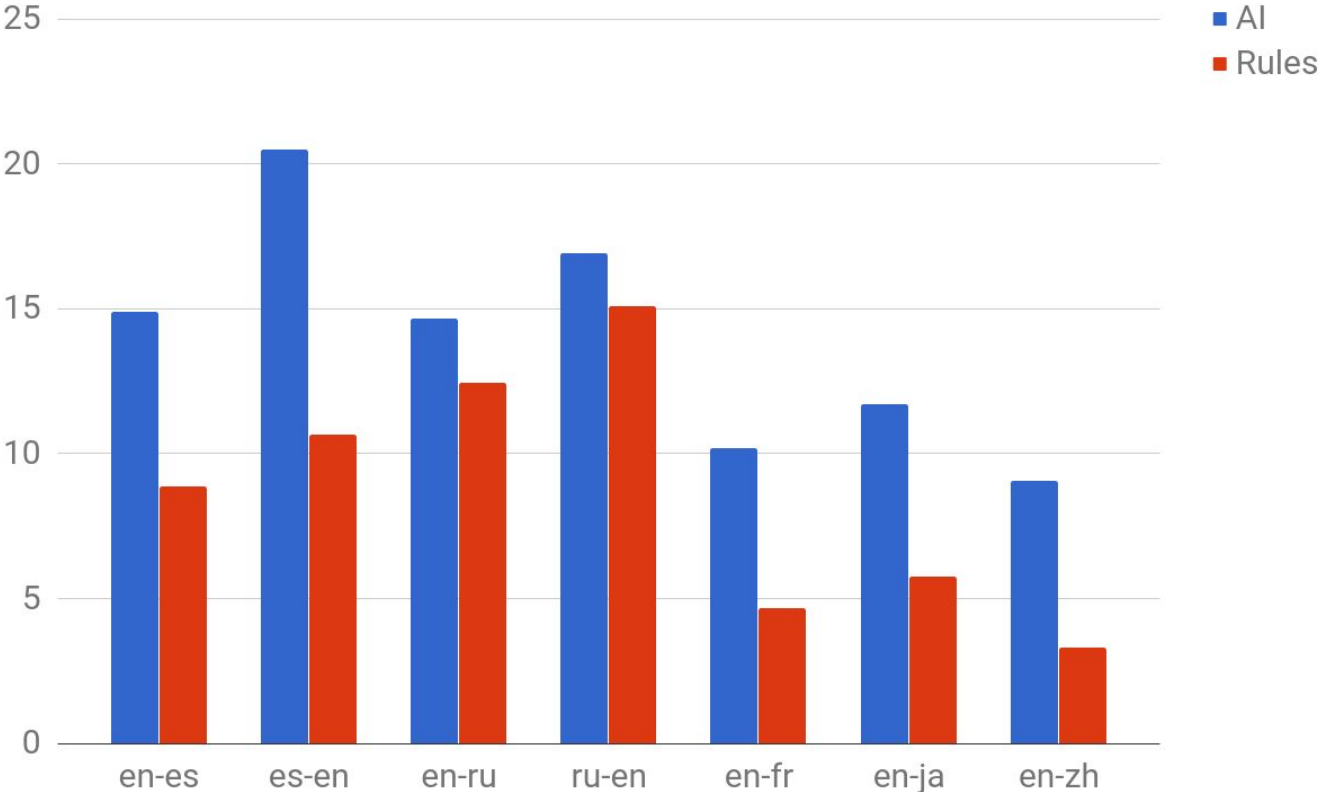
Filter Source Text Filter Target Text Aa Clear

#	Source: en	Target: cs			
1	16	16	✓	100	
2	16 mm	16 mm	✓	100	
3	john.doe@gmail.com	john.doe@gmail.com	✗	99	
4	Madrid	Madrid	✗	99	
5	Go to Madrid.		✗		
6	http://www.memsource.com/	http://www.memsource.com/	✗	99	
7	new_user_id	new_user_id	✗	99	
8	<xref:System.Linq.ParallelEnumerable.WithMergeOptions%2A>	<xref:System.Linq.ParallelEnumerable.WithMergeOptions%2A>	✗	99	
9	Spálená 51, 11000 Praha	Spálená 51, 11000 Praha	✗	99	

AI-BASED NON-TRANSLATABLE DETECTION

- Train on historical data
 - Which segments were changed, which stayed the same?
- Best performing approach: deep neural networks
- Be conservative
 - Avoid false non-translatables
- Each language pair is different
 - We support over 200 language pairs
 - Most of them have a separate model

EFFECT OF AI ON NT COVERAGE



NON-TRANSLATABLES: USER FEEDBACK

- Generally positive:
 - Savings
 - Allows to offer more competitive prices
- Complaints:
 - Inconsistency
 - False positives
- Based on production data, AI-based NT detection is several times more accurate than the rule-based solution.

OUTLINE

- Motivation: Enabling MT
- Identification of Non-Translatables
- **MT Quality Estimation**
- Next Steps
- Conclusion

MACHINE TRANSLATION IN CAT

- The most “obvious” machine-learning based feature
 - human translation reduced to post-editing MT outputs
- Neural MT can produce high-quality translations
 - conventional wisdom: NMT outputs are typically fluent and grammatical
 - ...but they can contain serious errors which are easy to miss
- Does it really save time?
 - ...that depends largely on MT quality
 - outputs which require no changes are still a minority

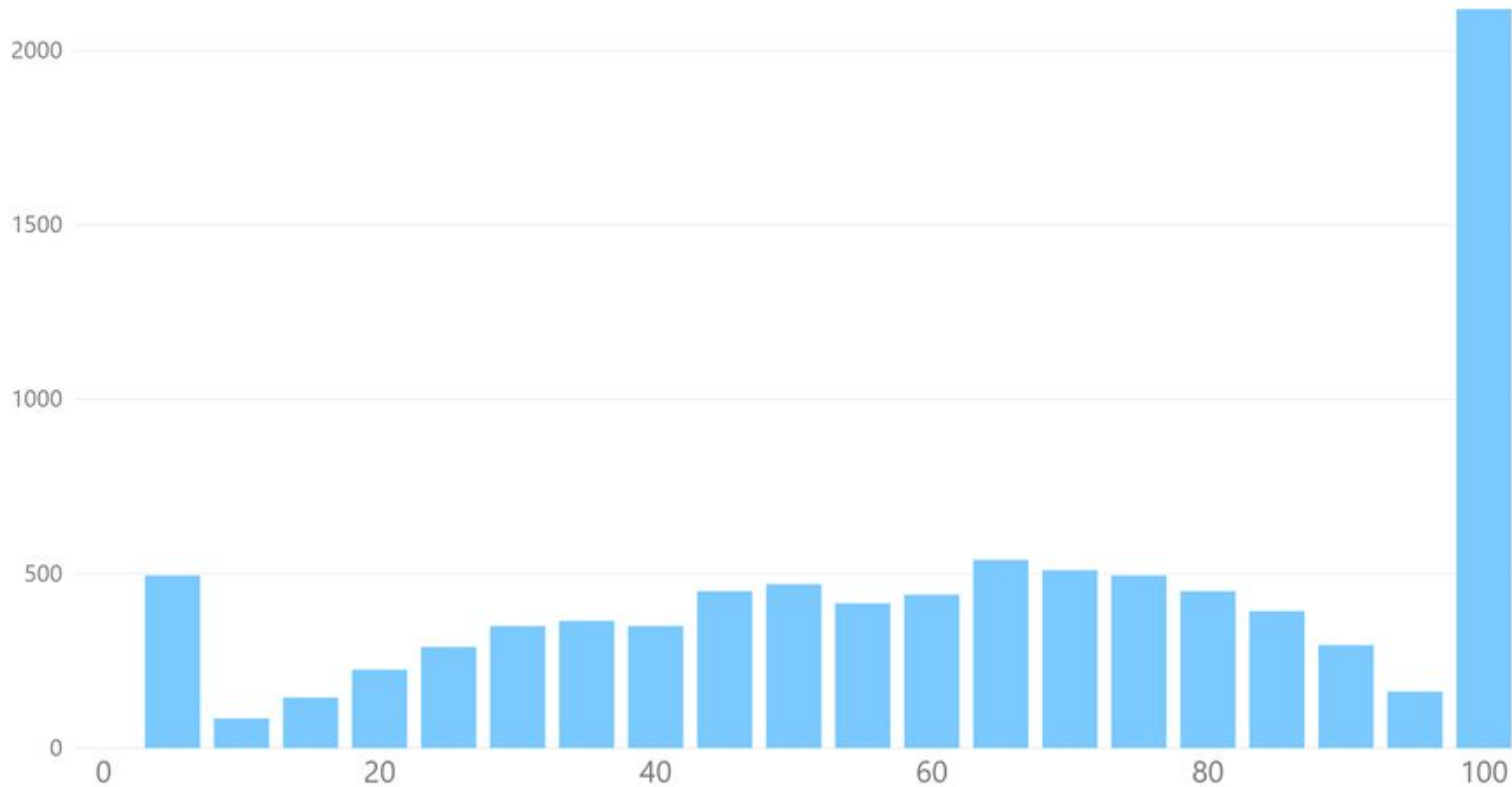
MT POST-EDITING: THE PROBLEM

How to identify the useful, high-quality MT outputs?

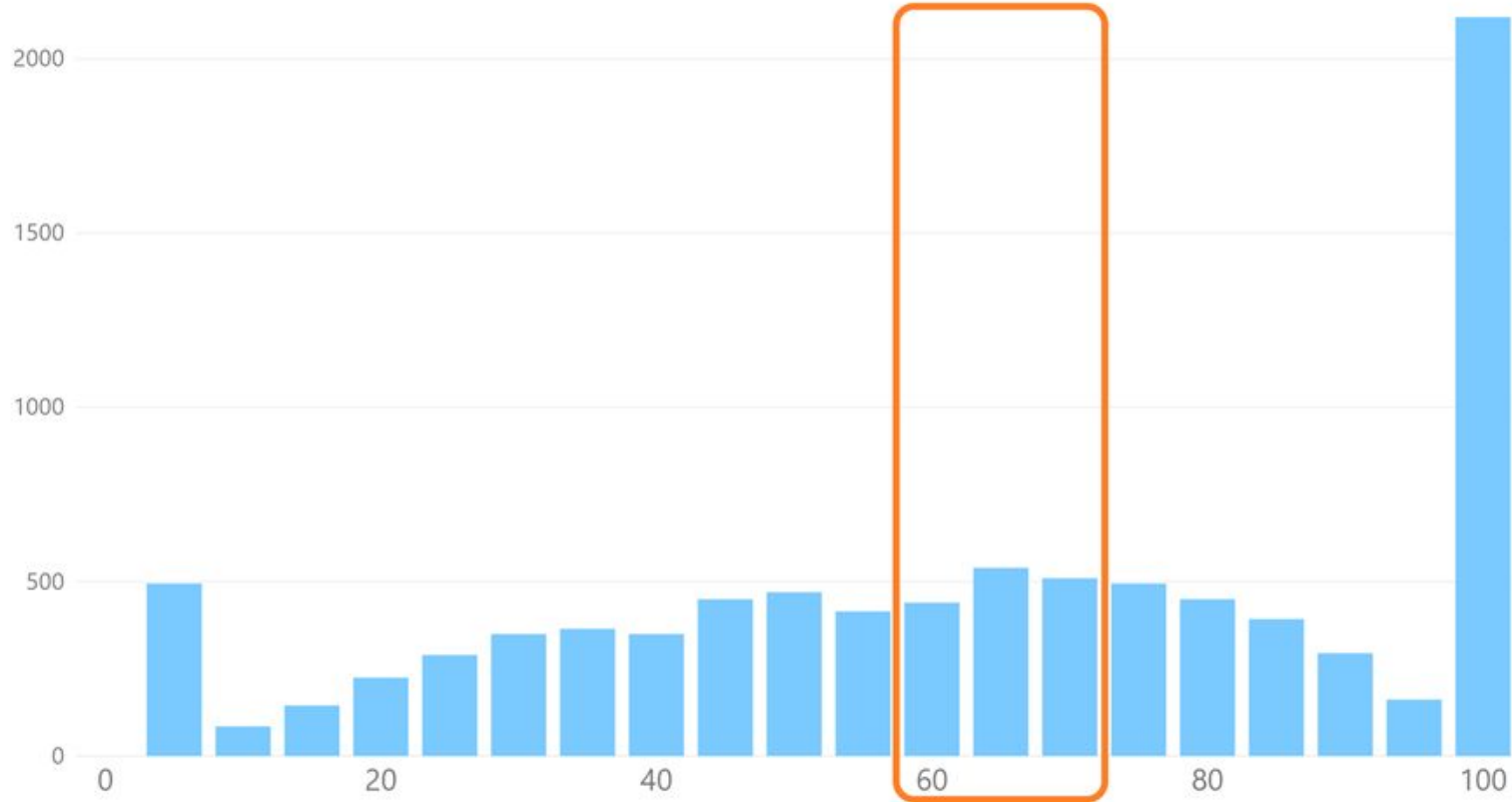
MTQE: USE CASES

- Decide whether to use MT for a project
- Predict translation cost ahead of time
- Score as a hint for translators

HOW DO WE MEASURE PERFORMANCE?

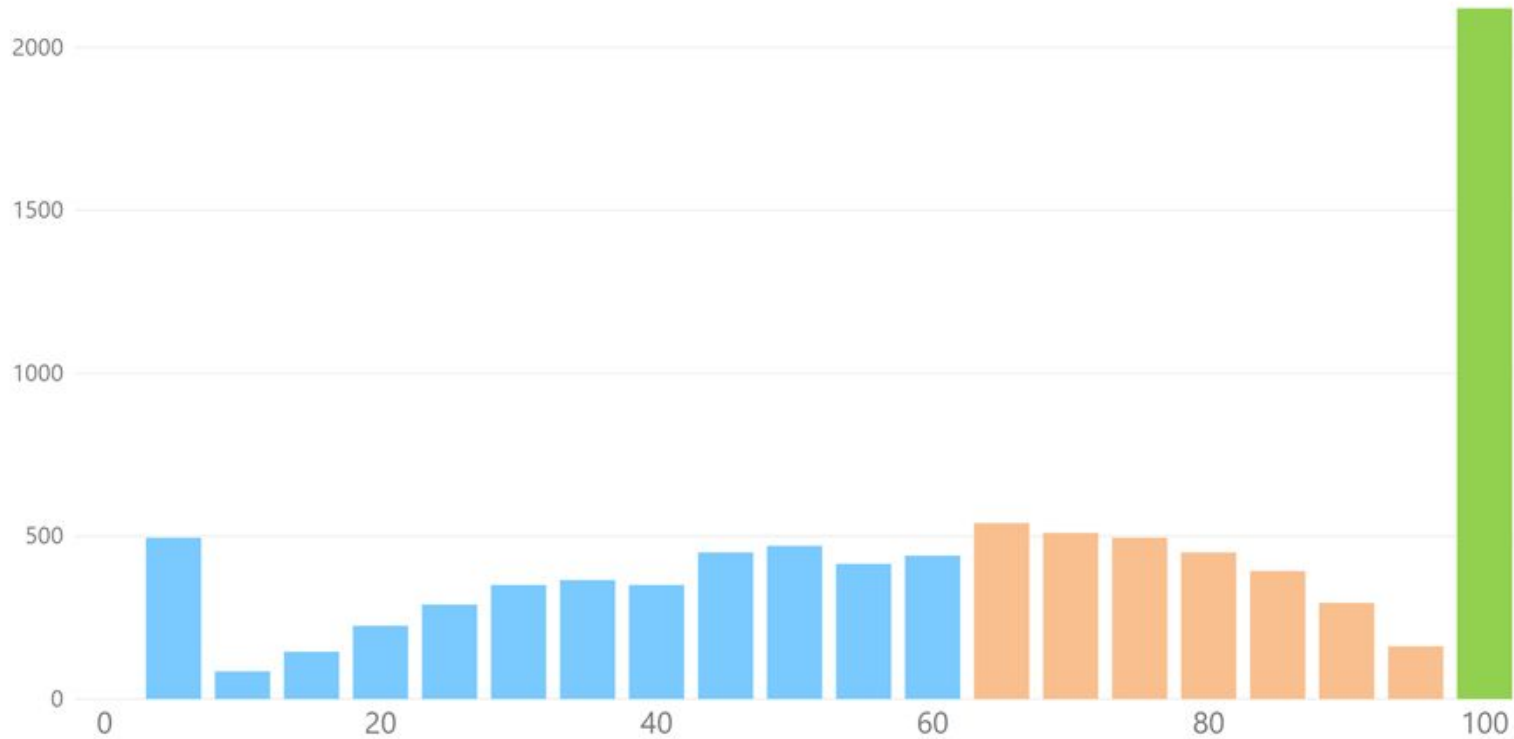


HOW DO WE MEASURE PERFORMANCE?



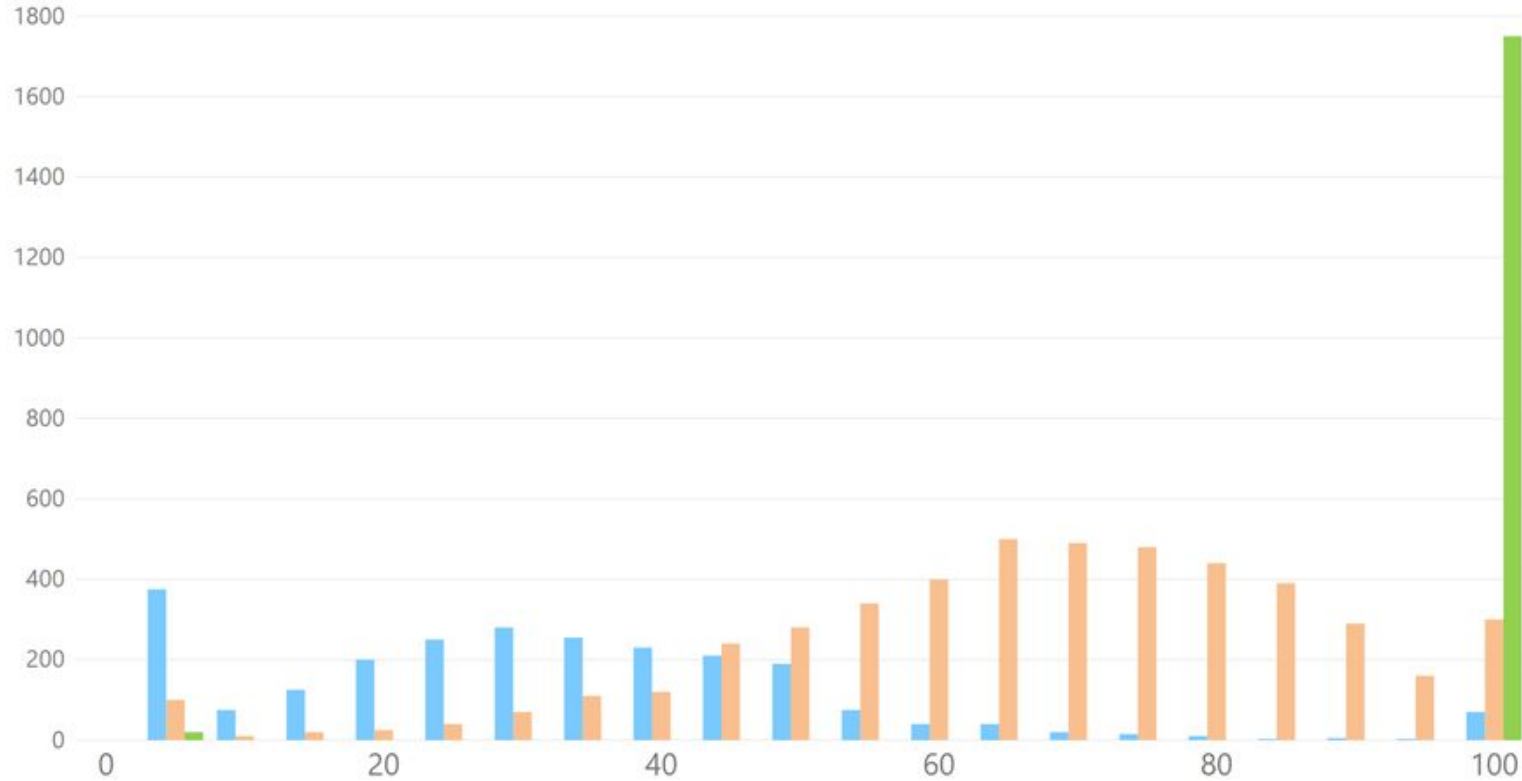
HOW DO WE MEASURE PERFORMANCE?

MTQE ● 0 ● 75 ● 100



HOW DO WE MEASURE PERFORMANCE?

MTQE ● 0 ● 75 ● 100



OUTLINE

- Motivation: Enabling MT
- Identification of Non-Translatables
- MT Quality Estimation
- **Next Steps**
- Conclusion

MEMSOURCE TRANSLATE: NEXT STEPS

- MT engine hub
 - Enables use of MT without complicated setup
 - Automatically selects MT engine based on language pair
 - Bundled with MTQE
- In the works:
 - Use **detected domain** in engine selection
 - **Continuous learning** of MT engine performance

NEXT STEPS: ADAPTIVE AND CUSTOM MODELS

- Adaptive MTQE models
 - MTQE predictions are adapted to the current document
- Custom MT engines in Memsource Translate
 - Automatically build and maintain MT engines tailored to the customer's content
 - Adaptive MT can be updated with new translations in real time

OUTLINE

- Motivation: Enabling MT
- Identification of Non-Translatables
- MT Quality Estimation
- Next Steps
- **Conclusion**

CONCLUSION

- MT can be difficult to use in translation industry
- Making MT integration seamless:
 - Automatic management of MT engines
 - Content-based MT engine selection
 - Analysis of MT performance ahead of translation
 - Segment-level score of MT quality
 - MT adaptation to new data

THANK YOU!

Q&A