

# Selection of MT Systems in Translation Workflows

Aleš Tamchyna  
October 8, 2020



# OUTLINE

- Introduction, motivation
- MT Quality Across Domains
- Approaches to MT Selection
- Conclusion

# INTRODUCTION

- MT quality has been steadily improving in the past few years
- MT can be very beneficial in translation
  - In some scenarios, MT can be used with little or no post-editing
  - MT can be a useful starting point for post-editing
- There are many commercial MT providers to choose from
  - Quality of MT systems varies across languages or domains
  - It is difficult to decide ahead of time which system is optimal for a project

## ABOUT MEMSOURCE

- Cloud-based translation management system
- Customers use Memsorce to manage the localization process and to produce translations
- We want to provide high-quality MT by default so that our users can benefit from MT as much as possible

# MT Quality Across Domains



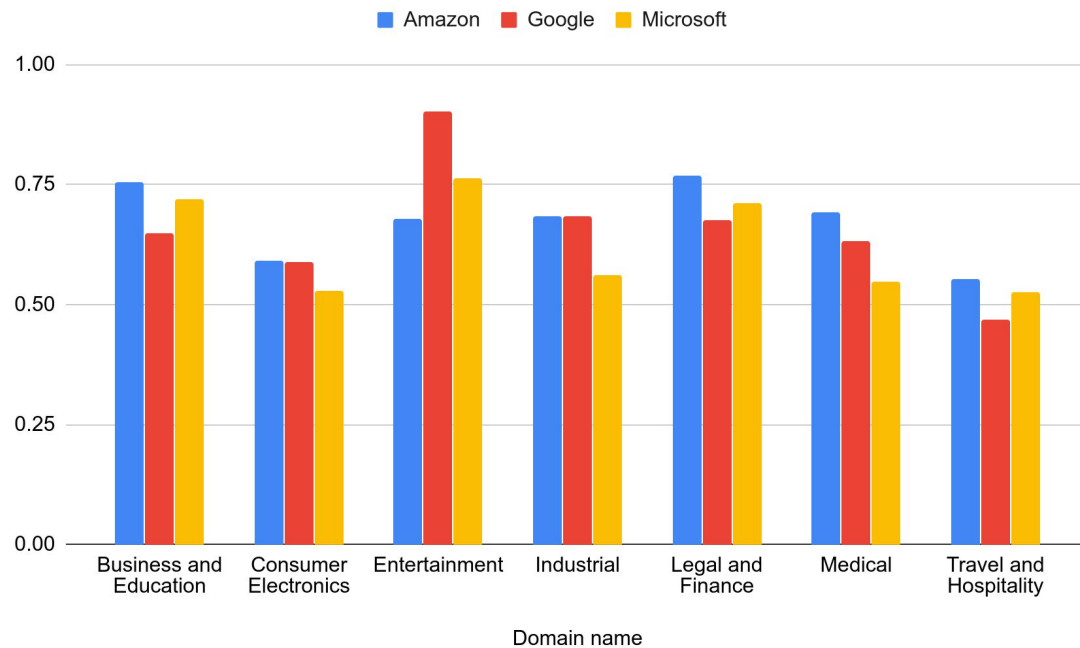
## METHODOLOGY

- Domains were defined using unsupervised machine learning on aggregate customer data, labels assigned manually
  - For non-English source languages, internal MT into English is applied first
- Domains contain data from multiple customers
- MT engines are assigned to documents using Memsorce Translate
  - Eliminates bias of customer preference for specific engines
  - Given enough data points, we can assume inputs for each MT system are i.i.d.

# DOMAINS

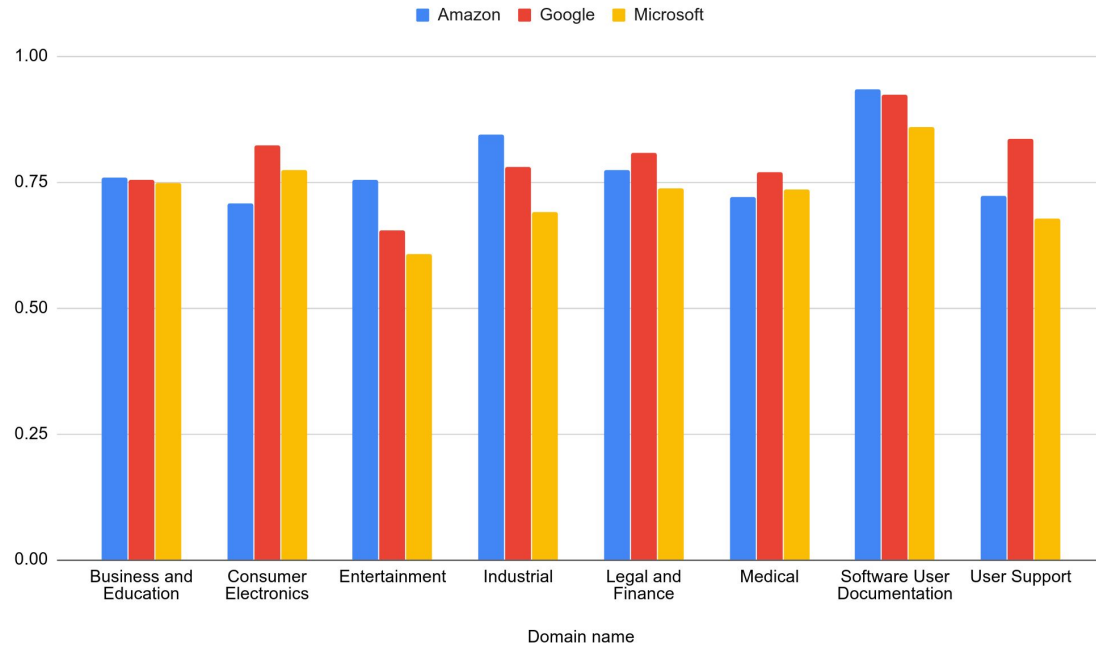
Domain	Keywords
Medical	'study', 'patients', 'patient', 'treatment', 'dose', 'mg', 'clinical'
Travel and Hospitality	'km', 'hotel', 'guests', 'room', 'accommodation'
Business and Education	'team', 'business', 'work', 'school', 'students',
Legal and Finance	'agreement', 'company', 'contract', 'services', 'financial'
Software User Documentation	'click', 'select', 'data', 'text', 'view', 'file',
Consumer Electronics	'power', 'battery', 'switch', 'sensor', 'usb',
User Support	'please', 'email', 'account', 'domain', 'contact',
Cloud Services	'network', 'server', 'database', 'sql', 'data'
Industrial	'mm', 'pressure', 'valve', 'machine', 'oil'
Software Development	'value', 'class', 'type', 'element', 'string'
Entertainment	'game', 'like', 'get', 'love', 'play', 'go', '

# RESULTS: ENGLISH-RUSSIAN



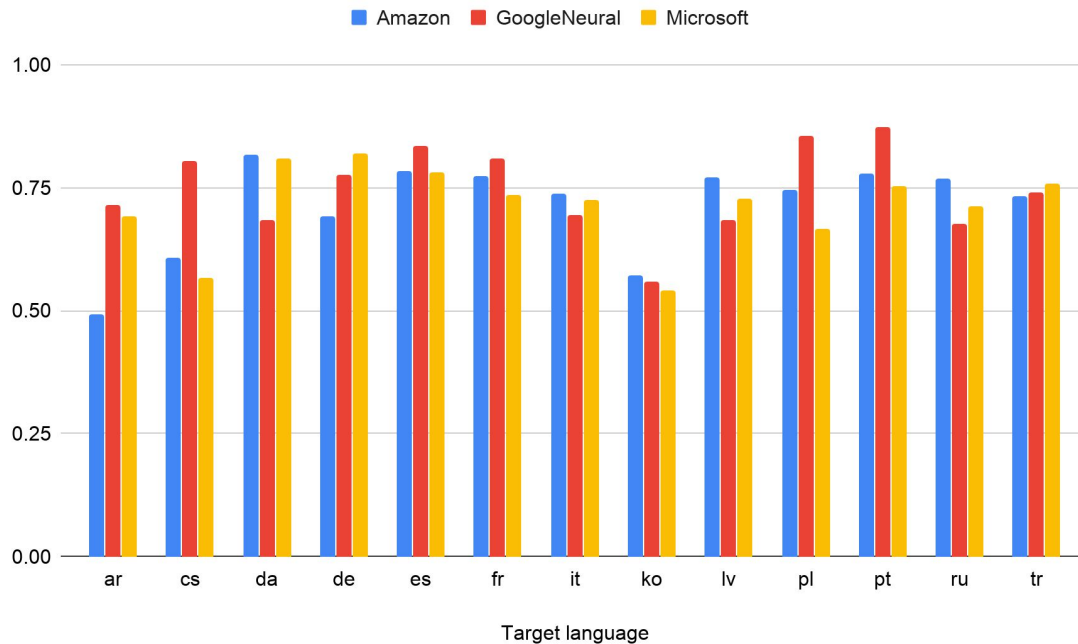


# RESULTS: ENGLISH-FRENCH



# RESULTS: SINGLE DOMAIN, MULTIPLE LANGUAGES

Domain: Legal and Finance



# IMPORTANCE OF SELECTING OPTIMAL MT ENGINES

- Given that all MT systems perform relatively well, does it matter which system is used?
- Sanchez-Torron and Koehn, 2016 show that *“for each 1-point increase in BLEU, there is a PE [post-editing] time decrease of 0.16 seconds per word, about 3-4%”*.
  - There is a clear correlation between MT quality and translator productivity.
  - The exact number may be different today due to specifics of NMT.

# Approaches to MT Selection



## PILOT STUDY

- High-level overview:
  - Create a sample dataset from the project
  - Translate the sample using multiple MT engines
  - Linguists are asked to post-edit the samples
  - Measure required amount of post-editing, time
- Robust, sound method but costly. Only makes sense for large projects/customers.
  - Needs to be re-done for every project (potential of data drift).

## MT QUALITY ESTIMATION

- Similar to pilot study but no manual post-editing.
- High-level overview:
  - Create a sample dataset from the project
  - Translate the sample using multiple MT engines
  - Measure MT quality using MTQE (manual translation is not required)
- Quick, cheap but still requires some manual steps (data preparation, evaluation).
  - Needs to be re-done for every project.
- MTQE may not be reliable enough for some domains/language combinations.

## MULTI-ENGINE MT

- Since multiple MT engines are available, use all of them.
  - MT system combination, not selection
- There are methods and for combining multiple MT outputs into a single translation, see e.g. Heafield and Lavie 2010, Freitag et al. 2014, Zhou et al. 2017
- More difficult to implement, costly (all engines used for all inputs), potentially the most robust option.

## MACHINE-LEARNING BASED SELECTION

- Use ML directly for recommending optimal MT engines based on translated content
- Only the selected MT engine is used (reduced costs)
- Fully automated for users, no manual steps are involved
- Commercial solutions:
  - Memsorce Translate
  - Smartling MT Auto Select
  - Intento Smart Routing\*
- Academic work is limited
  - At this conference though: Naradowsky et al. 2020, Machine Translation System Selection from Bandit Feedback

*\* It is not clear whether recommendations are based on ML or rather static benchmarks.*



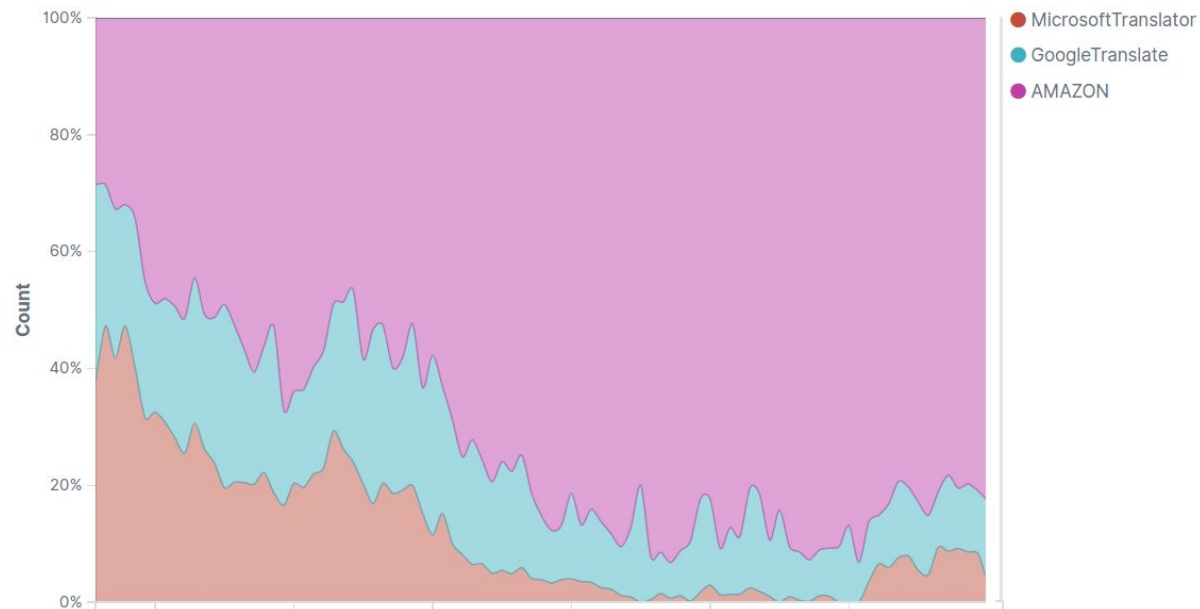
## MEMSOURCE TRANSLATE

- Automated selection of optimal MT system based on language pair and domain
- For every input document:
  - Analysis of content → domain label
  - Recommendation of MT system based on MT engine statistics
  - Once manual post-editing is completed, MT score is calculated → estimate update
- Recommendations driven by a standard algorithm for Bayesian multi-armed bandits
  - Model is continuously learning and improving
  - MT engine statistics are based on more than 100K documents (and growing)
- Simple, interpretable, fully automated
- A flexible framework, supports custom MT engines

# RECOMMENDED SYSTEMS IN TIME

English-Spanish, domain: User support

Recommendation by MT type



# Conclusion



# CONCLUSION

- MT can be very useful in localization
- Considerations:
  - Landscape of MT providers difficult to navigate
  - MT system quality varies across languages but also across domains
  - MT systems evolve over time
- Various approaches to MT selection exist
  - Manual evaluations work well for large, well-defined projects
  - Machine learning can allow to automate the process

## REFERENCES

- Axelrod, Amittai, Xiaodong He, and Jianfeng Gao. "Domain adaptation via pseudo in-domain data selection." In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 355-362. 2011.
- Freitag, Markus, Matthias Huck, and Hermann Ney. "Jane: Open source machine translation system combination." In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 29-32. 2014.
- Heafield, Kenneth, and Alon Lavie. "Combining machine translation output with open source: The Carnegie Mellon multi-engine machine translation scheme." *The Prague Bulletin of Mathematical Linguistics* 93, no. 2010 (2010): 27-36.
- Naradowsky, Jason, Xuan Zhang, and Kevin Duh. "Machine Translation System Selection from Bandit Feedback." *arXiv preprint arXiv:2002.09646* (2020).
- Sanchez-Torron, Marina, and Philipp Koehn. "Machine translation quality and post-editor productivity." In *Proceedings of AMTA*, vol. 1, pp. 16-26. 2016.
- Van der Wees, Marlies, Arianna Bisazza, Wouter Weerkamp, and Christof Monz. "What's in a domain? Analyzing genre and topic differences in statistical machine translation." In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 560-566. 2015.
- Zhou, Long, Wenpeng Hu, Jiajun Zhang, and Chengqing Zong. "Neural system combination for machine translation." *arXiv preprint arXiv:1704.06393* (2017).

# THANK YOU

# Q&A

