MT escaped from the Lab!
Now what?!

Mike Dillinger, PhD
Manager, Taxonomies & Machine Translation
LinkedIn

AMTA • Austin, TX • 2016
I work at the intersection of Machine Intelligence and Human Expertise

How can we integrate them?

- Machine Learning
- Human Translation
- Human Taxonomy Building
- Natural Language Understanding
- Text Comprehension
- Speech-to-Speech Translation
- Simultaneous Interpreting
- Knowledge Representation
- Domain Expertise
Congratulations!

Since 2010, use of MT at LSPs has grown dramatically.
Of small and large LSPs have at least a year’s experience with MT -- In 2016. In 2006, it was more like 5%.

MT really has escaped from the lab!
How did *that* happen?! 

Adopters of new technologies have a common decision making process (*Diffusion of Innovations*, Rogers, 1962) – Understand relative advantages over existing alternatives; Ease of observing advantages  

– Understand compatibility with existing processes  

– Lower complexity of deployment  

– Higher availability for testing

Much more translator education (Laurie Gerber, Jay Marciano, Rubén de la Fuente, me) since 2008. This yielded easier access to post-editors.

MUCH easier setup and training for non-technical users (Microsoft Hub, Google Translator Toolkit, Kantan, etc.) since 2010
Why is it important that MT escaped from the lab?

1. We continue to have jobs and get more resources :) 
2. Now Researchers are the main bottleneck for increased MT use; we can’t blame the translators anymore.
3. The demands and requirements for our systems are changing.

Red alert! We have new masters.
Sample Requirements: MT Users want to…

• Choose the best system for their content
• Manage terminology, named entities
• Predict grammatical quality
• Identify the worst errors
• Track quality, for workflow
• Customize systems surgically
• Improve locale-specific processing, tag processing / CAT-tool integration
• ...

WHY can’t we do these things yet?
MT Researchers are from Mars

We can do it alone!

MT Users are from Venus

We need help!

Because...
Researchers and Users have dramatically different **Goals**

MT in the lab: Build Autonomous Translation Machines
MT in practice:
Integrate MT with workflow and human expertise

Note: This is an accurate representation of a real translation workflow ; )
Researchers and Users focus on dramatically different **use cases** (1)

The *Clients* are very different.

<table>
<thead>
<tr>
<th>MT Researchers: <em>Inbound</em> Translation</th>
<th>Commercial MT Users: <em>Outbound</em> Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information consumption</td>
<td>Information dissemination, branding</td>
</tr>
<tr>
<td>Main client: Search engine</td>
<td>Main client: LSP, company</td>
</tr>
<tr>
<td>Many languages into English</td>
<td>English into many languages</td>
</tr>
<tr>
<td>Casual use, ephemeral content</td>
<td>Professional use, valuable content</td>
</tr>
<tr>
<td>Recall/Coverage is most important</td>
<td>Precision/Quality is most important</td>
</tr>
</tbody>
</table>

AMTA is the only venue where we can compare these use cases.
Researchers and Users focus on dramatically different **use cases** (2)

The *Priorities* are very different.

<table>
<thead>
<tr>
<th>MT Researchers: <em>Inbound</em> Translation</th>
<th>Commercial MT Users: <em>Outbound</em> Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall/Coverage is most important</td>
<td>Precision/Quality is most important</td>
</tr>
<tr>
<td>All domains are important</td>
<td>Only one domain is important</td>
</tr>
<tr>
<td>Train with all data available: More data = better data</td>
<td>Train only with relevant data: More data = more chaos</td>
</tr>
</tbody>
</table>

Here are a couple of examples...
Idiosyncrasies of training corpora are common...

<table>
<thead>
<tr>
<th>Source</th>
<th>MT output (Guess which training corpus we used!)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinkedIn members</td>
<td>LinkedIn member states of the European Union</td>
</tr>
<tr>
<td>Buy a gift for your mother</td>
<td>Buy a present for your motherf***er</td>
</tr>
<tr>
<td>This product comes with no</td>
<td>This product comes with guarantee.</td>
</tr>
<tr>
<td>guarantee.</td>
<td></td>
</tr>
<tr>
<td>Job title: Leader</td>
<td>Job title: Führer</td>
</tr>
<tr>
<td>A case for your iPhone</td>
<td>A lawsuit for your iPhone</td>
</tr>
</tbody>
</table>

More data = more chaos
Researchers and Users focus on dramatically different **use cases** (3)

The notions of *Quality* are very different.

<table>
<thead>
<tr>
<th>MT Researchers: <em>Inbound</em> Translation</th>
<th>Commercial MT Users: <em>Outbound</em> Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall/Coverage is most important</td>
<td>Precision/Quality is most important</td>
</tr>
<tr>
<td>Quality: best translation available</td>
<td>Quality: most useful translation</td>
</tr>
<tr>
<td>Quality goal: highest likelihood in (irrelevant) training corpus</td>
<td>Quality goals: most grammatical, most consistent terminology</td>
</tr>
<tr>
<td>Headaches: dramatically incorrect translations</td>
<td>Headaches: locale-specific conventions, tag management, inconsistent terminology</td>
</tr>
<tr>
<td>Quality measure: BLEU</td>
<td>Quality measure: Utility, throughput</td>
</tr>
</tbody>
</table>

High-likelihood translations are different from low-human-intervention translations.
Researchers and Users focus on dramatically different **use cases** (4)

*System Management* is very different.

<table>
<thead>
<tr>
<th>MT Researchers: <em>Inbound</em> Translation</th>
<th>Commercial MT Users: <em>Outbound</em> Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use my system, duh!</td>
<td>Which system to choose?</td>
</tr>
<tr>
<td></td>
<td>How to choose an MT system?</td>
</tr>
<tr>
<td>Improve the system: hyper-parameters</td>
<td>No access to hyper-parameters</td>
</tr>
<tr>
<td>Improve the system: re-train</td>
<td>What are the best new sentences to use?</td>
</tr>
<tr>
<td>Improve the system: heuristics</td>
<td>No access to internal heuristics</td>
</tr>
<tr>
<td>Improve terminology use?</td>
<td>Improve terminology use!</td>
</tr>
<tr>
<td>Let’s do a little domain adaptation</td>
<td>How well will this system work for my new content? What do I need to change? How much effort will it take?</td>
</tr>
</tbody>
</table>

In the end, humans have to use and manage MT systems.

©Mike Dillinger, 2016
MT Researchers are from Mars

MT Users are from Venus

Sometimes, we seem to be living in parallel universes!
Neural MT is hypnotizing MT researchers...

But will it help MT users?

Not if our use cases continue to be so different.
MT escaped from the lab.
Now what?

Neural MT is making you re-think your tools.

Now I want you to re-think your goals, to meet the needs of commercial MT use cases.
We’ve been thinking in terms of Machine only vs Human only translation for a long time. This is exceptionally simplistic, black vs white reasoning.

Our next challenge is to explore the 50 shades of grey in the middle.
Change One Assumption

Old Assumption:
Human translators do not have to be involved in translation. Machines can do it alone.

New Assumption:
Human translators have to be involved in translation, especially for commercial MT use cases.
Gedankenexperiment

Old top goal:
Build autonomous translation machines. Humans can fix the mistakes if they want to.

This focus gave us huge advances in methods:
– Fast development of new systems
– Lots of parallel data
– Probabilistic inference
Gedankenexperiment

New top goal:

Build hybrid-intelligence translation environments.
Leverage the different things that humans and machines do best. Assume multiple agents and enable collaboration. Think of AI to scale translation. Machine-only and human-only translation are just the edge cases.

Key hyper-parameter: % participation of humans

How would MT research change?
New Areas of Research

• Quality and Precision; Consistency tracking
  – Measure what humans and machines do best
  – Which humans do better than others for project X?
    Which machines do better than others for project X?
  – High-precision detection of what machines and
    humans do unreliably (~ confidence scoring)

• Hand-off mechanisms from human to machine
  control and vice versa: when and how?

We know from fly-by-wire systems for pilots
that these topics are very important
New Areas of Research

Enable system > translator communication:
- Detect and display translation issue types and density
- Identify and track top terminology, top collocations
- Are translators and machines all translating consistently?
- Provide document and project-level analytics

Enable translator > system communication:
- Uptake feedback from corrections (like e.g., Lilt does)
- Update feedback from rules
- Uptake feedback from exemplary translations

Communication is the heart of collaboration.
New Areas of Research (3)

• High-precision MT engines
• Better system combination: Choosing the best hypotheses for project X
• Better dynamic domain adaptation for fewer, more flexible systems to maintain
• Paraphrase detection and source re-writing for translation candidate generation
• Etc.

Technology to help humans get the job done.
In sum...

This is what we have:

Autonomous Translation Machines

This is what we need
Action Items for Hybrid-Intelligence Translation

1. Understand what humans and machines each contribute to translation, building on research into post-editing and the cognition of translation

2. Enable richer communication between humans and machines, building on research into adaptive MT and data visualization

3. Make MT systems more flexible and controllable, building on work in domain adaptation, system combination, and Neural MT

And keep going from there...
Because two heads are better than one!

Thanks for your attention!