An introduction to Machine Translation

Mike Dillinger, PhD
Principal, Translation Optimization Partners
mike@translationOptimization.com

Overview

• The problem: industrial-scale translation

• FAQs: what’s MT?
  • Can machines really translate?! 
  • Can we fire our translators now?

• Limitations: what MT can’t do
  • Why is the output so bad? What is MT good for?

• Return on investment: Benefits of MT
  • How much does MT cost?
  • How can we convince our bosses to buy MT?

• Workflow: MT in action
  • Why buy MT if it’s free on the internet?
  • What other kinds of translation automation are there?
  • How do we use it?

• Kinds of MT Systems
  • Rule-based, statistical, and hybrid
The Problem: Industrial-scale translation

Houston, we have a problem.

Communication is the lifeblood of business. Without communication, business can’t happen.
- Communication with clients
- Internal communication

But now our clients and operations are global

It’s too hard, too time consuming, and too expensive to re-write everything from scratch in each language.

So, translation is inescapable.
And it’s not just 20 or 30 pages, eh?
Why MT?

We need *industrial-scale* translation, part 1

There are more products = *more content*

The products are more complex = *more content*

The products have more uses = *more content*

Products change more rapidly = *more content*

Operations are more complex = *more content*

Content is used in more places = *more translation*

We need *industrial-scale* translation, part 2

Product development is faster = *faster* translation

Time to market is faster = *faster* translation

Support has to be faster = *faster* translation

Organizations have to be flexible = *faster* translation

We need *industrial-scale* translation, part 3

Oh, and by the way,

You can’t spend any more money to get all of this done! = *cheaper* translation
"Industrial scale" means...

More Better Faster Cheaper

More Better Faster Cheaper

More Better Faster Cheaper

More Better Faster Cheaper

A Perfect Storm

The practice of global communication is falling apart:

a) Goals
   Publish documents
   60% or more are not used

b) Scale
   Easier access to more documents; language pairs
   Information overload; crisis of confidence; a few languages

c) Process
   write + translate + use + support are all independent
   Each interferes with the other
The problems

**Scalability, cost, time**
- Human translation is (usually) wonderful but it doesn’t scale well
  - Bigger projects = more costs
  - Bigger projects = more issues
  - More languages = more costs + more issues
- Human translation is expensive
- Human translation is slow

Goals

We need **industrial-scale** translation processes: more, better, faster, cheaper

Introduction to MT

**FAQs:** What’s MT?
What's MT?

Machine Translation systems are software products that translate electronic texts (and speech) into other languages automatically.

• Do you mean systems like Google Translate and Babelfish?
  • Yes, the basic technology is the same, but for companies we adapt the system extensively, to meet your needs. The result is very different!

• Can we fire our human translators?
  • No. In most situations, MT requires human translators. Their job just changes so they can do more translation faster. Many translation agencies already use MT for draft translations because it saves them time and money.

• We already use machine translation from Trados, right?
  • Trados is one good use of old machine translation technology – it’s called “translation memory”. It doesn’t work well with new sentences or new topics. Modern machine translation technology can do a much better job with new input; think of MT as “translation reasoning”.

• You guys really hate translators, don’t you?
  • Not at all! Some overly enthusiastic MT researchers in the old days talked about replacing humans, which scared the pants off the translators. It also embarrassed us to death. Nowadays, we even invite translators to our conferences :)

• MT is ridiculous. Only humans can really translate. You have to understand the subtleties of language and culture.
  • It turns out that very, very many kinds of sentences are routine enough that machines can do a great job without subtle understanding. Most useful texts are neither poetic nor sophisticated.

• I’ve seen MT on the web. It’s laughable junk.
  • Millions of people use MT every day and very few complain. Besides, it’s free. What would a free Mercedes look like? Brand new Enterprise MT, customized to your needs is very, very different from what you see on the web.
FAQs: What’s MT?

“MT will have a negative impact on my brand”

Your site translated without MT

Your site translated with MT

Faster
Cheaper
More consistent

FAQs: Who’s really using MT?

- “In-bound” Translation (from other languages to yours)
  - Global Public Health Information Network (Public Health, Canada)
  - Many military and business organizations
  - Internet users around the world

- “Out-bound” Translation (from yours to other languages)
  - Symantec, Adobe, Cisco, Microsoft, Intel, European Community, etc.
  - Internet users around the world

- “Real-time” Translation (between two languages)
  - Translated subtitles (news, Jay Leno), translated TV and radio broadcasts
  - Internet users around the world: translated chat, translated SMS
Limitations: What MT can’t do

Introduction to MT

Limitations: What MT can’t do

Translation “quality”

Quality of target document

f (quality of source document + quality of target sentences)

So far, we’ve only used translators’ criteria for quality.

What do end users notice and not notice? What are their criteria for quality?

Correct source/target equivalence is still a question of trust, not of measurement.

- Should translators “fix” errors in the source documents?
- Should they reorganize source documents?

Dimensions of information quality:

• Content quality (relevant, complete, accurate information)
• Design quality (easy to find and maintain information)
• Linguistic quality (easy to understand information)
  • Term consistency
  • Stylistic simplicity
• Process quality (cost, consistency, reliability, etc.)
Limitations: What MT can’t do

How good?

Which MT system should I choose?

How not to evaluate MT

The usual (mis)steps:

• Hear salesperson say how great product X is
• Ask lots of questions about “quality”, speed, interfaces, cost
• Get an evaluation versions of product X and others
• Translate some of your documents
• Ask translators about “quality” of translation
• Get puzzled about poor output quality
• Decide not to use MT

Outcomes

• Wasted time, effort, money
• Little understanding, little learning
• Negative reputation for MT

Conclusion:

“These MT products are junk!”
Garbage in, garbage out?

Terminology Management

Dictionary Customization

Standard File Types

File Format Filters

Style Management

Grammar Customization

Conclusion:
This software rots!

Limitations:
What MT can't do

Tout le monde tente de sauter sur cette aventure, a déclaré Nicholas Field, qui permet de gérer environ 11 milliards de dollars en actions des marchés emergents a Schroders Plc à Londres.

“Everyone is trying to jump on that bandwagon,” said Nicholas Field, who helps manage about $11 billion in emerging-market stocks at Schroders Plc in London.
Tout le monde essaie de prendre ce train en marche, a dit Nicolas Field qui aide pour diriger approximativement $11 milliard dans les actions des marchés emergents à Schroder Plc à Londres. "Everyone is trying to jump on that bandwagon," said Nicholas Field, who helps to manage about $11 billion in emerging-market stocks at Schroders Plc in London.

The MSCI Emerging Markets Index rose 35 percent. This beat a 2.9 percent advance in the MSCI World Index of developed economies and lifted the value of stocks to $8.6 trillion from $5.1 trillion in 2008.

China’s part surpassed $3 trillion yesterday for the first time since August, from $1.8 trillion at the end of 2008.

The 22 nations that were classified as "emerging countries" by index provider MSCI Inc. comprised 24 percent of world market capitalization. This increased from 18 percent at the start of this year. This was the highest proportion since Bloomberg began compiling the data in 2003.

China's part surpassed $3 trillion yesterday for the first time since August, from $1.8 trillion at the end of 2008.

The increase signals growing confidence in developing countries as equity investors. Spurred by interest-rate cuts and stimulus plans, they are redeploying cash after the worst U.S. losses since the Great Depression.

The MSCI Emerging Markets Index rose 35 percent. This beat a 2.9 percent advance in the MSCI World Index of developed economies and lifted the value of stocks to $8.6 trillion from $5.1 trillion in 2008.

"Everyone is trying to jump on that bandwagon," said Nicholas Field, who helps to manage about $11 billion in emerging-market stocks at Schroders Plc in London.

The 22 nations that were classified as "emerging countries" by index provider MSCI Inc. comprised 24 percent of world market capitalization. This increased from 18 percent at the start of this year. This was the highest proportion since Bloomberg began compiling the data in 2003.

Les 22 nations qui ont été classées comme "pays émergents" par fournisseur de l'indice MSCI Inc. comprennent 24 pour cent de capitalisation du marché mondial. Cela a augmenté de 18 pour cent au début de cette année. C'était la plus haute proportion depuis que Bloomberg a commencé à compiler les données en 2003.

La partie de Chine a dépassé $3 trillion hier pour la première fois depuis août, de $1.8 trillion à la fin de 2008.

The increase signals growing confidence in developing countries as equity investors. Spurred by interest-rate cuts and stimulus plans, they are redeploying cash after the worst U.S. losses since the Great Depression.

The MSCI Emerging Markets Index rose 35 percent. This beat a 2.9 percent advance in the MSCI World Index of developed economies and lifted the value of stocks to $8.6 trillion from $5.1 trillion in 2008.

"Everyone is trying to jump on that bandwagon," said Nicholas Field, who helps to manage about $11 billion in emerging-market stocks at Schroders Plc in London.

The 22 nations that were classified as "emerging countries" by index provider MSCI Inc. comprised 24 percent of world market capitalization. This increased from 18 percent at the start of this year. This was the highest proportion since Bloomberg began compiling the data in 2003.

Les 22 nations qui ont été classées comme "pays émergents" par fournisseur de l'indice MSCI Inc. comprennent 24 pour cent de capitalisation du marché mondial. Cela a augmenté de 18 pour cent au début de cette année. C'était la plus haute proportion depuis que Bloomberg a commencé à compiler les données en 2003.

La partie de Chine a dépassé $3 trillion hier pour la première fois depuis août, de $1.8 trillion à la fin de 2008.

The increase signals growing confidence in developing countries as equity investors. Spurred by interest-rate cuts and stimulus plans, they are redeploying cash after the worst U.S. losses since the Great Depression.

The MSCI Emerging Markets Index rose 35 percent. This beat a 2.9 percent advance in the MSCI World Index of developed economies and lifted the value of stocks to $8.6 trillion from $5.1 trillion in 2008.

"Everyone is trying to jump on that bandwagon," said Nicholas Field, who helps to manage about $11 billion in emerging-market stocks at Schroders Plc in London.

The 22 nations that were classified as "emerging countries" by index provider MSCI Inc. comprised 24 percent of world market capitalization. This increased from 18 percent at the start of this year. This was the highest proportion since Bloomberg began compiling the data in 2003.

Les 22 nations qui ont été classées comme "pays émergents" par fournisseur de l'indice MSCI Inc. comprennent 24 pour cent de capitalisation du marché mondial. Cela a augmenté de 18 pour cent au début de cette année. C'était la plus haute proportion depuis que Bloomberg a commencé à compiler les données en 2003.

La partie de Chine a dépassé $3 trillion hier pour la première fois depuis août, de $1.8 trillion à la fin de 2008.

The increase signals growing confidence in developing countries as equity investors. Spurred by interest-rate cuts and stimulus plans, they are redeploying cash after the worst U.S. losses since the Great Depression.

The MSCI Emerging Markets Index rose 35 percent. This beat a 2.9 percent advance in the MSCI World Index of developed economies and lifted the value of stocks to $8.6 trillion from $5.1 trillion in 2008.

"Everyone is trying to jump on that bandwagon," said Nicholas Field, who helps to manage about $11 billion in emerging-market stocks at Schroders Plc in London.

The 22 nations that were classified as "emerging countries" by index provider MSCI Inc. comprised 24 percent of world market capitalization. This increased from 18 percent at the start of this year. This was the highest proportion since Bloomberg began compiling the data in 2003.

Les 22 nations qui ont été classées comme "pays émergents" par fournisseur de l'indice MSCI Inc. comprennent 24 pour cent de capitalisation du marché mondial. Cela a augmenté de 18 pour cent au début de cette année. C'était la plus haute proportion depuis que Bloomberg a commencé à compiler les données en 2003.

La partie de Chine a dépassé $3 trillion hier pour la première fois depuis août, de $1.8 trillion à la fin de 2008.

The increase signals growing confidence in developing countries as equity investors. Spurred by interest-rate cuts and stimulus plans, they are redeploying cash after the worst U.S. losses since the Great Depression.

The MSCI Emerging Markets Index rose 35 percent. This beat a 2.9 percent advance in the MSCI World Index of developed economies and lifted the value of stocks to $8.6 trillion from $5.1 trillion in 2008.
 Lessons learned

- Technology is designed and built for optimal performance in specific conditions (ex: paved road, competent driver, correct fuel)

  Even a wonderful, brand-new BMW looks like junk when it’s tested outside the design specs.

- Using technology outside of its “comfort zone” requires adaptation.

Action items

- Understand MT’s “comfort zone”
- Assess which adaptations are necessary:
  - Input
  - People
  - Process
  - Technology

The “comfort zone” for MT

<table>
<thead>
<tr>
<th>Adaptations by writer</th>
<th>Adaptations of MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manage terminology and vocabulary explicitly</td>
<td>Familiar to the system</td>
</tr>
<tr>
<td>Manage writing style explicitly</td>
<td>Words and phrases</td>
</tr>
<tr>
<td>Manage writing style explicitly</td>
<td>Sentence types</td>
</tr>
<tr>
<td>Use standard file formats</td>
<td>Fluent, predictable meanings</td>
</tr>
<tr>
<td>Write to minimize post-editing</td>
<td>Standardized file formats</td>
</tr>
<tr>
<td></td>
<td>Post-editing</td>
</tr>
</tbody>
</table>

- Very fast processing
- Very large volumes
- Good quality

Limitations: What MT can’t do

- What MT can’t do
The status quo

- Ill-defined authoring processes
- Simple localization management processes
- Informal work flow definition
- Few quantitative metrics in use

Why deploying MT seems difficult

- Term detection and management
- Well-defined authoring processes
- Systematic quality control of source
- Document pre-processing
- Post-editing
- Systematic quality control of target
- Error analysis for feedback
- Document post-processing

Who can do this?
- In-house staff?
- New hires?
- Outsourced?
Introduction to MT

ROI: Benefits of MT

Machine Translation systems provide faster and cheaper translations than humans with translation memory tools alone.

- MT captures translator knowledge and effort in additional ways (memory vs. reasoning)
- MT requires more disciplined writing, which leads to additional efficiency and savings
- MT shifts the translator’s workload from slower, more complex tasks (translation) to faster, simpler tasks (revising)

There are a range of different scenarios for translation automation, with and without MT.
**ROI:** Benefits of MT

**Total effort (time * cost)** for the same 1,000,000-word project

Area = amount of effort

<table>
<thead>
<tr>
<th>Cost (USD)</th>
<th>Delivery time (person-days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$250,000</td>
<td>20</td>
</tr>
<tr>
<td>$240,000</td>
<td>20</td>
</tr>
<tr>
<td>$230,000</td>
<td>20</td>
</tr>
<tr>
<td>$220,000</td>
<td>20</td>
</tr>
<tr>
<td>$210,000</td>
<td>20</td>
</tr>
<tr>
<td>$200,000</td>
<td>20</td>
</tr>
<tr>
<td>$190,000</td>
<td>20</td>
</tr>
<tr>
<td>$180,000</td>
<td>20</td>
</tr>
<tr>
<td>$170,000</td>
<td>20</td>
</tr>
<tr>
<td>$160,000</td>
<td>20</td>
</tr>
<tr>
<td>$150,000</td>
<td>20</td>
</tr>
<tr>
<td>$140,000</td>
<td>20</td>
</tr>
<tr>
<td>$130,000</td>
<td>20</td>
</tr>
<tr>
<td>$120,000</td>
<td>20</td>
</tr>
<tr>
<td>$110,000</td>
<td>20</td>
</tr>
<tr>
<td>$100,000</td>
<td>20</td>
</tr>
<tr>
<td>$90,000</td>
<td>20</td>
</tr>
<tr>
<td>$80,000</td>
<td>20</td>
</tr>
<tr>
<td>$70,000</td>
<td>20</td>
</tr>
<tr>
<td>$60,000</td>
<td>20</td>
</tr>
<tr>
<td>$50,000</td>
<td>20</td>
</tr>
<tr>
<td>$40,000</td>
<td>20</td>
</tr>
<tr>
<td>$30,000</td>
<td>20</td>
</tr>
<tr>
<td>$20,000</td>
<td>20</td>
</tr>
<tr>
<td>$10,000</td>
<td>20</td>
</tr>
<tr>
<td>$1,000</td>
<td>20</td>
</tr>
<tr>
<td>$0</td>
<td>20</td>
</tr>
</tbody>
</table>

**Benefits of MT**

<table>
<thead>
<tr>
<th>Time</th>
<th>Internal benefits</th>
<th>Market benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Delivery time 4 times faster or more</td>
<td>More sales opportunities</td>
</tr>
<tr>
<td></td>
<td>More flexible</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shorter launch schedule</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scarcity</td>
<td>Better user experience</td>
</tr>
<tr>
<td></td>
<td>More consistent terminology use</td>
<td>Better indexing and search</td>
</tr>
<tr>
<td></td>
<td>Better indexing and search</td>
<td>Better user experience</td>
</tr>
<tr>
<td></td>
<td>Lower operating costs</td>
<td>More funds for improvements</td>
</tr>
<tr>
<td></td>
<td>Generally 50% lower</td>
<td>Better user experience</td>
</tr>
<tr>
<td></td>
<td>More languages</td>
<td>Better user experience</td>
</tr>
<tr>
<td></td>
<td>Less translation effort per language</td>
<td>Scalability</td>
</tr>
</tbody>
</table>
How does Translation Automation save time and money?

**Task analysis**

Translation includes different activities, each with different speeds and costs.

**Translators:**
- **Translate** from scratch other sentences (non-matches)
- **Revise** translations that are worth fixing. ("fuzzy" matches)
- **Approve** translations that are correct. ("perfect" matches)
- **Skip** sentences that have already been translated. ("ICE" matches)

Different tools divide "translation" into these activities in different ways.

**Important assumption:**
Output quality is the same in all scenarios.

**Scenario 1**

<table>
<thead>
<tr>
<th>Sample project</th>
<th>Sample project 1,000,000 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>17¢/wd</td>
</tr>
<tr>
<td>25%</td>
<td>10¢/wd</td>
</tr>
<tr>
<td>15%</td>
<td>3¢/wd</td>
</tr>
<tr>
<td>10%</td>
<td>1¢/wd</td>
</tr>
</tbody>
</table>

**Total time:** 293 person-days  
**Total cost:** USD $115,963
**Translation Workflow Scenarios**

<table>
<thead>
<tr>
<th>Words 1,000,000</th>
<th>Translate with no tools</th>
<th>TM only</th>
<th>Customized MT only</th>
<th>TM + MT with existing TM</th>
<th>TM + MT with extra customization</th>
<th>TM + MT with Better source</th>
<th>Raw MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% of project</td>
<td>Translate</td>
<td>Revise</td>
<td>Approve</td>
<td>Translate</td>
<td>Revise</td>
<td>Approve</td>
<td>Translate</td>
</tr>
<tr>
<td>Total cost (USD)</td>
<td>$170,000</td>
<td>$115,963</td>
<td>$92,670</td>
<td>$81,453</td>
<td>$69,923</td>
<td>$61,213</td>
<td>$43,280</td>
</tr>
<tr>
<td>Cost per word (USD)</td>
<td>0.17</td>
<td>0.13</td>
<td>0.09</td>
<td>0.08</td>
<td>0.07</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>% change</td>
<td>-33%</td>
<td>-19%</td>
<td>-13%</td>
<td>-14%</td>
<td>-12%</td>
<td>-30%</td>
<td>-28%</td>
</tr>
</tbody>
</table>

**A focus on price?**

1,000,000 words

<table>
<thead>
<tr>
<th>Cost per word</th>
<th>Total cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.30</td>
<td>$300,000</td>
</tr>
<tr>
<td>0.25</td>
<td>$250,000</td>
</tr>
<tr>
<td>0.20</td>
<td>$200,000</td>
</tr>
<tr>
<td>0.15</td>
<td>$150,000</td>
</tr>
<tr>
<td>0.10</td>
<td>$100,000</td>
</tr>
<tr>
<td>0.05</td>
<td>$50,000</td>
</tr>
<tr>
<td>0.00</td>
<td>$10,000</td>
</tr>
<tr>
<td>0.01</td>
<td>$1,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Delivery time (days)</th>
<th>Lost sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>$273,973</td>
</tr>
<tr>
<td>30</td>
<td>$821,918</td>
</tr>
<tr>
<td>60</td>
<td>$1,643,836</td>
</tr>
<tr>
<td>90</td>
<td>$2,465,753</td>
</tr>
<tr>
<td>120</td>
<td>$3,287,671</td>
</tr>
<tr>
<td>150</td>
<td>$4,109,589</td>
</tr>
<tr>
<td>180</td>
<td>$4,931,507</td>
</tr>
<tr>
<td>7</td>
<td>$191,781</td>
</tr>
</tbody>
</table>

Or on delivery time?

Focusing on a lower vendor price per word is a recipe for disaster.
How much does MT cost?

Evaluation system

**Direct costs**
- Trial installation: US$1,000 for desktop version (or loaner from vendor)
- Consultant for planning and training personnel

**Indirect costs**
- Personnel time

Production system

**Direct costs** (initial installation)
- Server: US$30,000 - $200,000 per language pair
- Vendor or consultant services for MT customization/training
- Consultant for planning and training personnel

**Indirect costs**
- Maintenance fee: ~20% of server price, per year
- Personnel time

What are these costs for? (1)

<table>
<thead>
<tr>
<th>Input</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitor translatibility</td>
<td>Beneficial without MT</td>
</tr>
<tr>
<td>Convert to standard file formats</td>
<td>Beneficial without MT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Writers: Manage terminology and vocabulary explicitly</td>
<td>Beneficial without MT</td>
</tr>
<tr>
<td>Writers: Manage writing style explicitly</td>
<td>Beneficial without MT</td>
</tr>
<tr>
<td>Writers: Use standard file formats</td>
<td>Beneficial without MT</td>
</tr>
<tr>
<td>Editors: Train to edit for MT input</td>
<td>Beneficial without MT</td>
</tr>
<tr>
<td>Project Managers: Train to minimize post-editing and delivery time</td>
<td>Beneficial without MT</td>
</tr>
<tr>
<td><strong>MT operator</strong>: Train or hire</td>
<td>***</td>
</tr>
</tbody>
</table>
What are these costs for? (2)

<table>
<thead>
<tr>
<th>Process</th>
<th>Technology</th>
</tr>
</thead>
</table>
| Develop pre- and post-processing tools and procedures | Add words and phrases to dictionary | Beneficial without MT  
| Develop evaluation metrics | Extend grammatical coverage; couple translation memory |  
|  | Extend semantic coverage; couple translation memory | ***  
|  | Add filters and converters | ***  
|  | Extend system performance to minimize post-editing | ***  
|  | Integration with existing systems | ***

Benefits of MT

Faster, cheaper translation:
• Better scalability
• Better capture and re-use of translator knowledge and effort
  - Complements TM
• Shifts translator workload to simpler tasks
• Promotes better writing

Discussion

Questions?
**Workflow:** MT in action

- **Introduction to MT**

**More output from each translator**

- Word processor: Investment in tools: none
- Translation Memory: Investment in tools: small
- **TM + MT**: Investment in tools: large

Industrial-scale translation
More output from each translator

**One** translator action yields:

- Translation of 1 sentence, in one place, in one document, into one language: 1:1

- Translation of some source sentences, in different places, in many documents, into many languages: 1:1,000

- Translation of all source sentences, in many documents, into many languages: 1:10,000

---

**Leveraging one action into many**

<table>
<thead>
<tr>
<th>Translator’s action</th>
<th>Result</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual Translation</td>
<td>One translated segment, in one place, in one document</td>
<td>None</td>
</tr>
<tr>
<td>Translate one segment in one document</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We need to capture and multiply translator effort = create "leverage"

<table>
<thead>
<tr>
<th>Next place? Next document?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translate one segment in one document</td>
</tr>
<tr>
<td><strong>With Translation Memory</strong></td>
</tr>
<tr>
<td>The whole translated segment in different places, in many documents</td>
</tr>
<tr>
<td>- Not &quot;portable&quot; to different domains</td>
</tr>
<tr>
<td>- Some leverage, some revision</td>
</tr>
<tr>
<td>- Skip some sentences</td>
</tr>
<tr>
<td>- Revise more</td>
</tr>
<tr>
<td>- Translate from scratch less</td>
</tr>
</tbody>
</table>

| Translate one segment in one document |
| **With TM + Machine Translation** |
| Pieces of the same translated segment in different places, in many documents |
| - Now "portable" to different domains |
| - More leverage, more revision |
| - Skip more sentences |
| - Revise more |
| - Translate from scratch less |
Focus

More from each translator

Focus on multiplying the results of translators’ effort:
• Capturing effort is important
• Re-using effort is important
• Different tools have different limitations

• Translation tools capture and re-use translator effort to improve scalability
  • In different ways

Discussion

Questions so far?

The Content Supply Chain

Write
  Plan Content
  Gather/Review Information
  Draft/Revise Text

Accuracy?
Completeness?

Relevance?

Tech Writers

Accuracy?
Completeness?
Readability?
Translatability?

From Translators? End Users? Support? QA?

No feedback = no improvement
The Content Supply Chain

Translate

Filter with TM

Draft Translations

Revise Translations

Translators

MT Engines

Glossary?

Translation Review?

Marketing?

Sales?

End Users?

Support?

QA?

Fight with:
- file formats
- formatting
- version control
- graphics

No feedback, No improvement

Improve MT:
- Dictionary Customization
- Grammar Customization

No feedback = no improvement

Use Information

Find information

Understand information

Apply information

End Users

Usefulness?

Accuracy?

Completeness?

Relevance?

Easy to find?

Accuracy?

Completeness?

Easy to understand?

Usefulness?

No feedback = no improvement

The Content Supply Chain
Wait a second!
I can just use Google Translate (or Babblefish, Bing Translator, etc.)! That'll save me lots of money.
: ) : )
Right?

For lack of know-how, most organizations try to deploy MT...
The wrong way: MT as a “silver bullet”

Issues:
- No adaptation of source writing to MT limitations
- No explicit terminology management
- No on-going MT optimization
- No systematic re-use of feedback for error avoidance
- Massive post-editing is expected to compensate for poor implementation
**The right way: Step 1. Optimize processes without MT**

**Approach:**
- Create infrastructure for on-going optimization
- Accumulate know-how
- Use feedback and communication to prevent future errors

**The right way: Step 2. Add MT**

**Approach:**
- MT accelerates existing effective processes
- MT does not make up for lack of effective processes
- Optimization know-how is the competitive advantage
Assess adaptations that are needed

Input
• Make writing more translatable
• Standardize file formats

People
• Train writers
• Train/hire post-editors
• Train/hire MT operator(s)

Process
• Develop pre- and post-processing tools
• Develop metrics

Technology
• Customize/train MT

Action items
• Assess adaptations in more detail
• Estimate deployment effort

Kinds of MT systems: rule-based, statistical, and hybrid
### Kinds of MT systems

#### Rule-based MT

<table>
<thead>
<tr>
<th>Rule-based MT</th>
<th>Why you should care</th>
</tr>
</thead>
<tbody>
<tr>
<td>~600 words per second</td>
<td>Usually not a factor for localization</td>
</tr>
<tr>
<td>Better with word order</td>
<td>✔ Fewer complex edits</td>
</tr>
<tr>
<td>Better with sentence structure</td>
<td>✔ Fewer complex edits</td>
</tr>
<tr>
<td>Issues choosing phrasing and stylistics</td>
<td>☒ More edits about word choice</td>
</tr>
<tr>
<td>Targeted customization</td>
<td>✔ Can fix very specific errors and prepare the system for specific projects</td>
</tr>
<tr>
<td>Many tools for targeted customization</td>
<td>✔ Can fix very specific errors and prepare the system for specific projects</td>
</tr>
<tr>
<td>More complex to customize from existing translations</td>
<td>☒ On-going investment in system improvement</td>
</tr>
</tbody>
</table>

- Hard to build for new languages
- Generally less expensive

#### Statistical MT

<table>
<thead>
<tr>
<th>Statistical MT</th>
<th>Why you should care</th>
</tr>
</thead>
<tbody>
<tr>
<td>~200 words per second</td>
<td>Usually not a factor for localization</td>
</tr>
<tr>
<td>Issues with word order</td>
<td>☐ More complex edits</td>
</tr>
<tr>
<td>Issues with sentence structure</td>
<td>☐ More complex edits</td>
</tr>
<tr>
<td>Better at choosing phrasing and stylistics</td>
<td>☒ Fewer edits about word choice</td>
</tr>
<tr>
<td>Global customization</td>
<td>Very convenient but</td>
</tr>
<tr>
<td>Few tools for targeted customization</td>
<td>☐ Hard to make specific changes</td>
</tr>
<tr>
<td>Simple, efficient training from existing translations</td>
<td>☒ Very convenient built-in feedback to reuse human translations</td>
</tr>
</tbody>
</table>

- Easy to build for new languages
- For the moment, more expensive
- But only if you have many existing translations
### Kinds of MT systems

<table>
<thead>
<tr>
<th>Rule-based MT</th>
<th>Hybrid MT</th>
<th>Statistical MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>~600 words per second</td>
<td>~200 words per second</td>
<td></td>
</tr>
<tr>
<td>Better with word order</td>
<td>Issues with word order</td>
<td></td>
</tr>
<tr>
<td>Better with sentence structure</td>
<td>Issues with sentence structure</td>
<td></td>
</tr>
<tr>
<td>Issues choosing phrasing and stylistics</td>
<td>Better choosing phrasing and stylistics</td>
<td></td>
</tr>
<tr>
<td>Targeted customization</td>
<td>Global customization</td>
<td></td>
</tr>
<tr>
<td>Many tools for targeted customization</td>
<td>Few tools for targeted customization</td>
<td></td>
</tr>
<tr>
<td>More complex customization from existing translations</td>
<td>Simple, efficient training from existing translations</td>
<td></td>
</tr>
</tbody>
</table>

All plug into different content management systems

<table>
<thead>
<tr>
<th>Hybrid MT</th>
<th>Statistical MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard to build for new languages</td>
<td>Easy to build for new languages</td>
</tr>
<tr>
<td>Generally less expensive</td>
<td>For the moment, more expensive</td>
</tr>
</tbody>
</table>

---

### Action Items

Many factors pressure us to localize more, better, faster, and cheaper

- Use tools to leverage **one** translator action into many, **many** changes in the translated output
- Use tools to emphasize cheaper, faster activities
- Use tools for cost reduction and increased throughput
- Use editing feedback to improve tools
- Every investment in more consistent, more readable source documents yields huge returns for localization

---

### Wrap-up

**Learn more about MT**

- Don’t go it alone – **hire a consultant** to help choose and deploy MT
- Educate **all** your stakeholders about MT, continuously
We are independent translation automation consultants who help you to:

- Improve your strategic decision making and planning to get it right
- Understand your current situation and effective paths to reach your goals
- Troubleshoot your existing processes and tools to solve your immediate problems

About us

Principal:
Mike Dillinger
mike@translationoptimization.com

Thanks for your attention.

Questions?

An Introduction to Machine Translation

Mike Dillinger, PhD
Principal, Translation Optimization Partners
mike@translationoptimization.com