



Making Machine Translation
work for You

Mike Dillinger, PhD
Translation Optimization Partners

- ▶ Moderated by: Jennifer Beaupre
 - ▶ Director of Marketing, acrolinx North America
 - ▶ Jennifer.Beaupre@acrolinx.com

- ▶ Presented by: Mike Dillinger PhD
 - ▶ Principal, Translation Optimization Partners
 - ▶ Past President and Vice President, Association for Machine Translation in the Americas
 - ▶ Mike.Dillinger@TranslationOptimization.com

[\[Presentation\]](#)

Making MT work for you

Mike Dillinger, PhD

Principal, Translation Optimization Partners
www.translationOptimization.com

Translation Optimization Partners

June 16, 2010

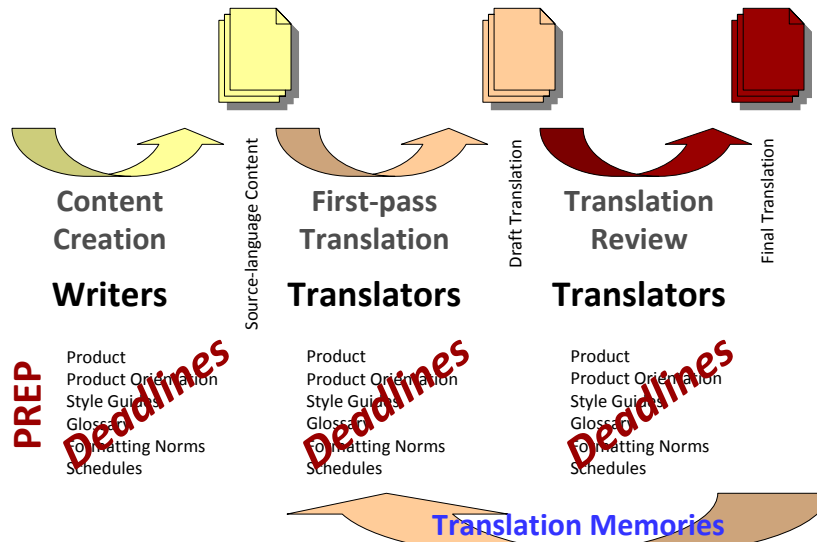
Cause > effect

MT output quality works by simple cause and effect:

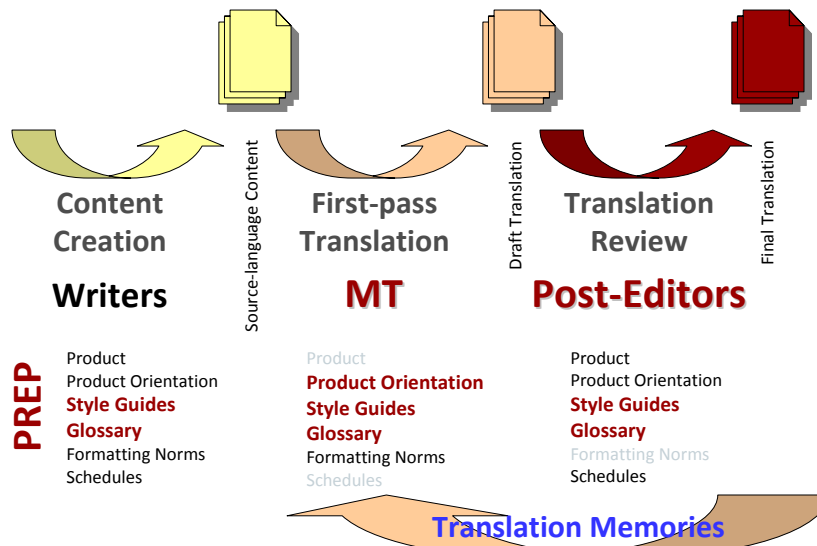
Your process and your content cause changes in MT quality.

When the causes are under your control, the effects are, too.

Without MT...



With MT...



Moral of the story

To make MT work for you,
build a systematic translation process:

- **Invest** (heavily) **in your Writers**
- **Invest** (continuously) **in your MT system**
- **Invest in your Post-Editors** (and listen to them)

Get help: These are unfamiliar, new processes and technologies.

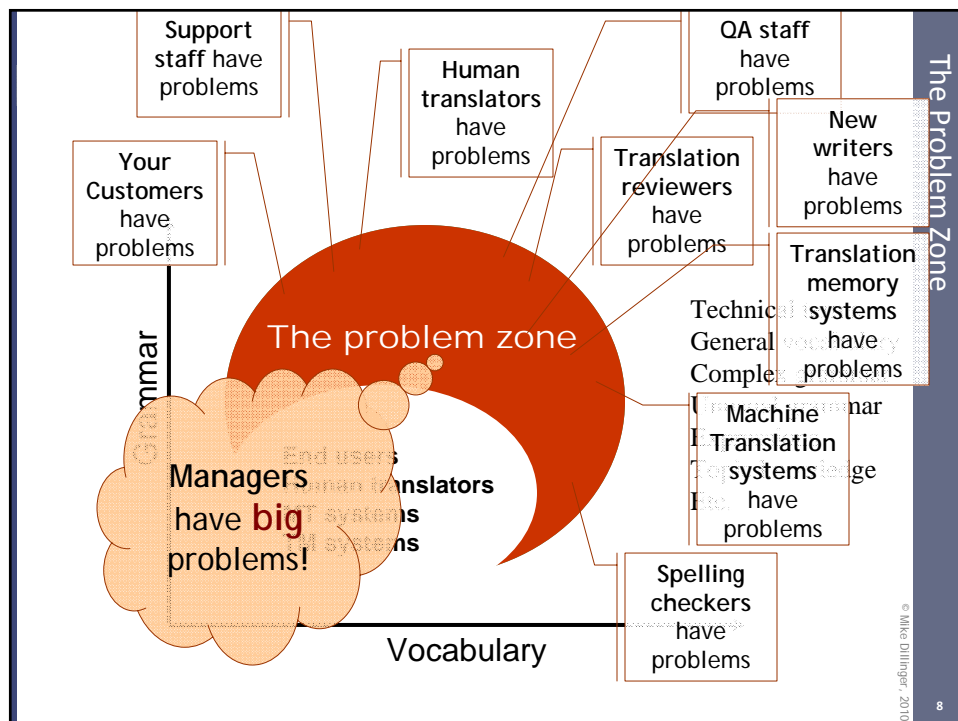
Invest (heavily) **in your writers**

To avoid self-inflicted translation costs

Writing for global audiences is a new skill

MT insider tip: Authors don't naturally write simply and clearly, i.e., appropriately for global audiences and for MT.

- They need systematic training in new skills;
- Normal writing style often produces poor MT output;
- Some choices for content management processes interfere with MT, as well;



What's clear to the writer...

These tend to be of a general nature and basically require employees to look after their own health and safety, and that of others, while working, to cooperate with employers in matters of health, safety and environmental protection, including properly using any equipment, procedures or protective measures provided, and not to behave recklessly.

One sentence:
53 words
27th grade reading level (!)
7 conjunctions
7 verbs

Even if the translation is perfect, will anyone understand?

From a Safety Manual!

Writing for MT?

The myth: MT requires special writing

The reality: *People* require special writing

Reuther (2003), O'Brien (2003)

The rules are pretty much the same for human translators and for machine translation systems [and for end users, as well]

Rules of writing for MT?

Rules shared by at least four controlled-language specifications (O'Brien, 2003):

Use only approved words from dictionary

- *When appropriate, use an article or demonstrative adjective before a noun*
- *Do not make noun clusters longer than three nouns*
- *Use noun phrases with fewer than 4 elements*
 - *Ex: Each route stop may have its own job plan , which the application copies to the **child route stop work order**.*
- *Avoid the gerund (-ing)*
- *Do not omit relative pronouns such as “who”, “which” or “that”*

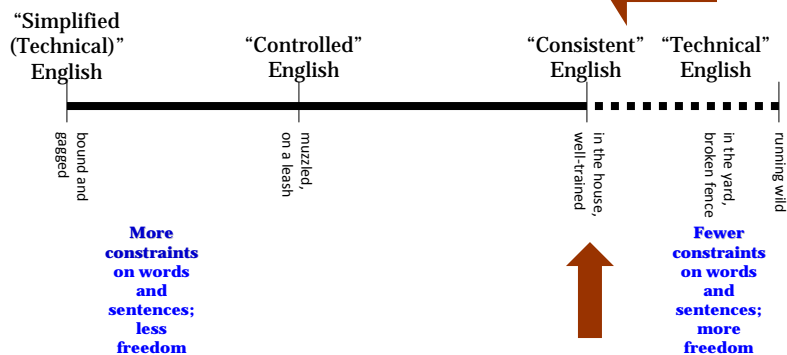
Use only active voice

Make instructions as specific as possible

This sounds very much like the Style Guide that you *already* use.

Don't write for MT; write for humans

Kinds of English



This is what TM and MT tools see

What Is #Brand #Product? In What Part of the Data Center Should #Brand #Product • Increases and fluctuations in demand: End-user

#Brand® #Product delivers the next step Be Deployed? demand for application performance is never

in the #Product vision by uniting network, Typical placement of #Brand #Product is consistent. IT organizations typically respond by

computing, and virtualization resources into a seamless in the access area of the data center. Connections should static overprovisioning, which wastes resources.

system. The system integrates a low-latency unified feed into a 10-Gbps aggregation switch such as the #Brand #Brand #Product can provision and reprovision

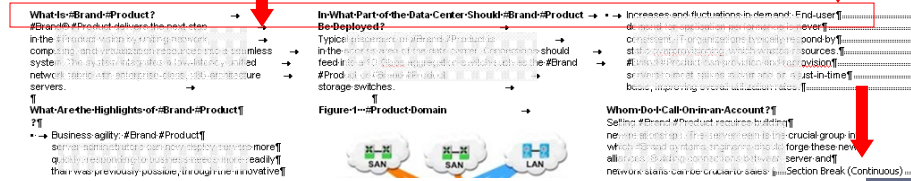
network fabric with enterprise-class, x86-architecture #Product or #Brand #Product servers to meet spikes in demand on a just-in-time

servers. storage switches. basis, improving overall utilization rates.

This is why

How-to-Sell #Brand-#Product

Selling-Guide



Humans are a real problem...

Cause > effect

Cause:

We don't have the time/money/support/know-how/tools/etc. to...

Optimize content reuse

But what happens when you don't?

20

© Mike Dillinger, 2010

17

Re-use your content

Topic reuse?

4. Type your user name and password. If this is the first time logging in, log in with the factory default username and password (admin and **admin** (all lower case)); you will be prompted to change the default password. Refer to Changing the Default Password for more information.

Setting a New Password

When you first log into the Remote Console, you are prompted to set a new password to replace the default. Then you can configure the **■■■**.

1. Log on to a workstation with network connectivity to your **■■■■■** unit.
2. Launch a supported Web browser such as Internet Explorer (IE) or Firefox.
3. In the address field of the browser, enter the default IP address of the unit:
■■■■■■■
4. Press Enter. The login page opens.
5. Enter the user name **admin** and the password **■■■■■**

Do you see the similarities in these two?

Duplicate effort for writers : (

And for translators...

© Mike Dillinger, 2010

18

Re-use your content

Sentence reuse?

Variety	Consistency
<p>18 After completing this module, the participant should be able to:</p> <p>16 After completing this module, you should be able to:</p> <p>15 After completing this module you should be able to:</p> <p>13 After completing this module the participant should be able to:</p> <p>7 By the end of this module, participants will:</p> <p>4 By the end of this module, participants should be able to:</p>	<p>79 After this module, you will be able to do all this:</p>
<p>Translation Cost: \$8.29</p>	<p>\$1.87</p>

Fluff

Fluff = wording that adds additional length but no additional information to a sentence.

Ex:

- *It is recommended that you disable Java...*
- *Disable Java...*

We systematically find **15%** of fluff in client content. Sometimes more.

- Fluff costs translators twice:
 - You pay for every word for each language, even when it's fluff.
 - Fluff interferes with matching previously-translated sentences.

Wordiness interferes with reuse

Port **has been** configured and connection to port is possible. (~400x)

Important: **It is recommended that you** disable Java caching and clear the Java cache. Please refer to your Java documentation or the **■■■■■** Multi-Platform Client (MPC) and **■■■■■** Remote Client (■■■) User Guide for more information.

(~16x)

The appropriate allowance **should be** made.
(~500x)

In this sample,
almost 1500 words = 6 pages

Parenthetical information [~6,275!]

A username can be deactivated (**Active checkbox is deselected when the number of failed login attempts has exceeded the maximum login attempts set in the Security Settings page**).

Cause > effect

Cause:

We don't have the time/money/support/know-how/tools/etc. to...

Get feedback from users

But what happens when you **don't**?

Relevance?

Microsoft

Not using feedback?

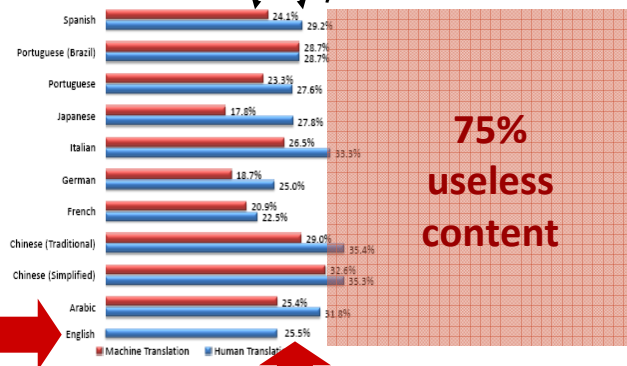
Provide feedback on this article

Did this article help you solve your problem?

- Yes
- No
- Partially
- I do not know yet

Customized MT is almost as useful as HT and is much cheaper

Knowledge Base – average resolve rate of human translated vs. machine translated articles



But...

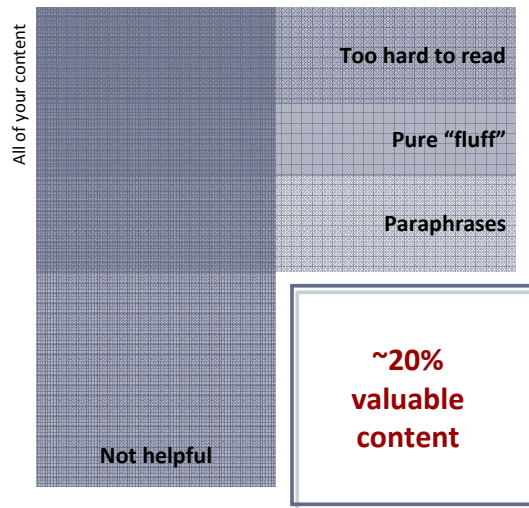
Wendt, 2008

© Mike Dillingner, 2010

23

Kinds of content you pay to translate

Control your costs at the source



You have *much* more control over your translation costs than any LSP does.

© Mike Dillingner, 2010

24

In practice...

Content QA: How global is *your* content?

- We do MT Readiness audits
- We do readability and translatability audits
- We do pre-translation editing
- We pinpoint problems and prioritize remediation
- We do style guide review
- We train writers

- Find and fix worst 20%
- On-going editing and monitoring
 - Avoid future issues

Invest (continuously) **in your MT system**

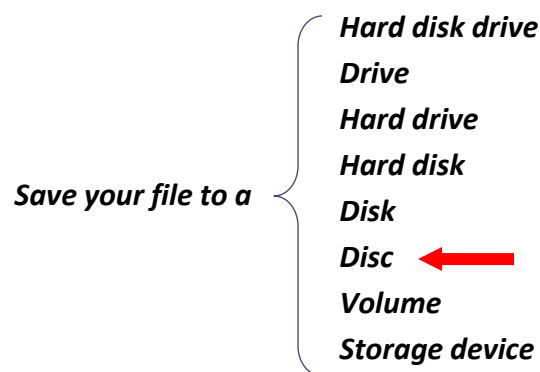
To reduce translation costs

MT is a toolkit

MT insider tip: A commercial MT system is a *toolkit*, not an “off-the-rack” product.

- It will cover *very* many core terms and sentence types in both languages;
- It will *not* cover all your writers’ “unique” ways of expressing themselves;
 - You need to adapt the toolkit to your specific style and vocabulary before starting to use it
(= *customization or training*; MT vendors will help)
 - You need to continuously update and adjust the system for best performance, as content changes
(= *linguistic maintenance or tuning*; MT vendors sometimes don’t help but offer tools)

Term reuse?



Synonyms make existing translations much harder to find and reuse.
So you pay to re-translate the same information.

Term reuse?

Which of these words do you really need?

They appeared only once in a 1,200,000-word doc set

abbreviation, accord, acknowledge, afterward, agreed, allegation, analysed, antecedent, antoher, anyhow, archives, assumptions, assures, audible, auditability, **averted**, beneficial, blanks, briefly, canadian, clearance, closer, compilation, complicated, **communicate**, concerns, conductors, conspicuously, contacted, **datagrams**, **denyed**, deployable, disclaimed, discontinue, elect, encountering, enjoy, exclusion, executions, expanding, exposed, **fashion**, feautres, fiche, fridays, functioning, gained, **gratis**, headed, **hyphen**, imaging, implicitly, **incomming**, incorporating, indefinitely, infinity, infringement, intact, integrators, intersection, intervention, isolation, kilo, leaflet, legal, licensing, linear, minimal, **minutres**, msec, neglect, numbering, optimally, originating, **periodaclly**, persistent, pipes, posted, privacy, programming, prohibited, publicize, **putput**, realize, replies, representative, reselect, resembles, resort, resuming, reveal, sale, school, **screwdriver**, secondly, segments, senses, shelf, simplicity, someday

75% of words and terms appear only once or twice in your content.
This makes MT more difficult, TM less effective, and reading harder.

Noun “stacks” ...

Returning User Group Information
System Settings Control Center
#Brand #Product Local Port Connection
#Product Manager Device System Information Configuration Backup
#Brand Communication Exchange Specifications
Automatic Color Calibration checkbox

Setting Sun Video Resolution
Setting Server Video Resolution

Cooking rocks! (a Rebecca Ray cookbook)
~how to cook rocks
~rocks for cooking
~cooking is cool

In practice...

MT Prep: How effective is your MT system?

- We plan MT deployments
- We help you choose the right MT system
- We do MT evaluation and troubleshooting
- We develop MT dictionaries
- We pinpoint problems and prioritize remediation

- ◀ Plan for MT
- ◀ Find and fix worst problems
- ◀ On-going monitoring

Invest in your Post-editors (and listen to them)

To improve the supply chain

Post-editing is a new skill

MT insider tip: Translators don't automatically understand how to post-edit MT output without specific training.

- They need systematic training;
- Post-editing forces us to define criteria for “good enough” translations

Post-editing is important

Ladies, leave your clothes here and spend the afternoon having a good time. (In a Rome laundry)

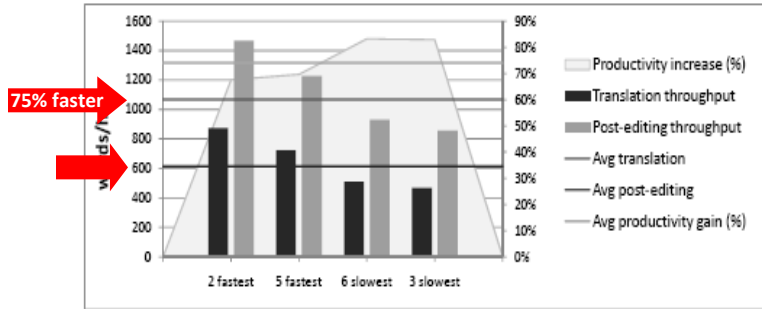
Ladies are requested not to have children in the bar. (In a Norwegian cocktail lounge)

Specialist in women and other diseases. (In the office of a Roman doctor)

Daily plate -- shrimp in spit. (In a Brazilian restaurant)

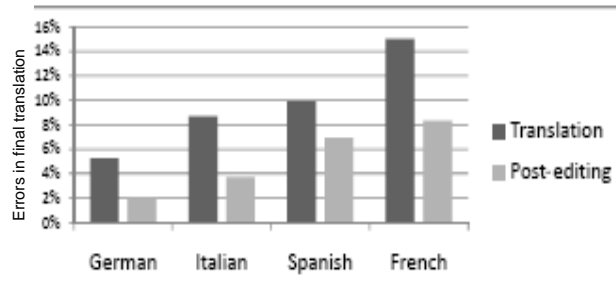
These are *human* translations, eh?

Post-editing is faster than translation



Plitt & Masselot, 2010

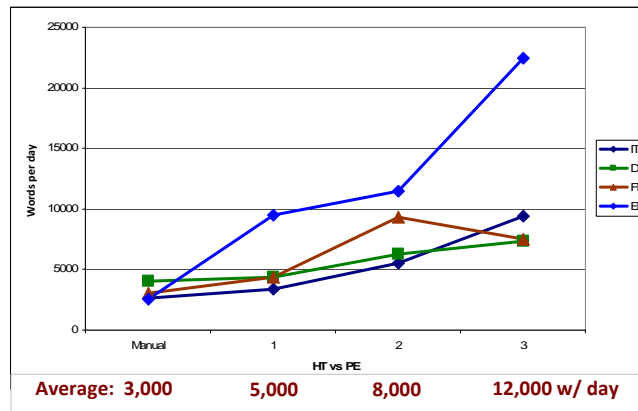
Post-editing is more accurate, too



Plitt & Masselot, 2010; see also Guerberof, 2009

Translators learn post-editing quickly

Even without improving the source...



Translator productivity during a 3-week post-editing course.

In practice...

Post-editing: How effective is your post-editing team?

- We train post-editors
- We evaluate and monitor post-editing performance
- We troubleshoot MT for error prevention

- Training
- On-going monitoring
- Fix worst problems

Next steps

To-do list:

1. **Assess your MT readiness today**
2. **Start training writers to communicate with global audiences**
3. **Build linguistic resources**
4. **Get help to develop your MT deployment plan**

Make MT work for you!

Audits & Edits & Plans

Translation Optimization Partners

Our Services

Pinpoint your missed opportunities

Source Content Audit

We pinpoint where & how, and then quantify just how much your source content interferes with translation, understanding, and branding.

Translation Process Audit

We pinpoint where & how, and then quantify just how much your current processes create self-inflicted costs.

Reduce self-inflicted costs and increase efficiencies

Pre-translation Editing

We optimize your source content for translation.

Author Training

We train your writers to optimize source content for translation.

Decision Support for Translation Automation

We guide you past the hype and confusion when you choose tools and vendors, plan process improvements, define requirements, develop RFPs, calculate ROI, identify risks.

Contact us at: mike@translationOptimization.com

© Mike Dillinger, 2010

41

Thanks for your attention.

Questions?

Translation Optimization Partners

Independent Consultants who help you to
Optimize content
Optimize translation
Optimize opportunities

Mike Dillinger, PhD

mike@translationOptimization.com

www.translationOptimization.com

- ▶ www.acrolinx.com
- ▶ Follow us on
 - ▶ www.twitter.com/acrolinx
 - ▶ www.facebook.com/acrolinx
 - ▶ www.linkedin.com – acrolinx, information IQ, aQuatic
- ▶ Upcoming Events- contact us for more information
 - ▶ Executive Forum- Boston, June 24, 2010
 - ▶ Webinar- The ROI of Information Quality with Gilbane, July 14, 2010
 - ▶ Executive Forum- Minneapolis, July 15, 2010