

Association for Machine Translation in the Americas

# AMTA - 2006 CONFERENCE

## TUTORIAL ON

### Tools and Techniques for Translatable Content

Presenter:  
Mike Dillinger

BOSTON MARRIOTT CAMBRIDGE  
CAMBRIDGE, MA

8 - 12 AUGUST 2006



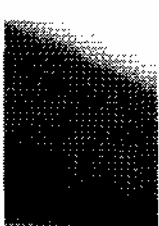
---

## *Tools and Techniques for Translatable Content*

Mike Dillinger, PhD  
mike@mikedillinger.com

Association for Machine Translation in the Americas, 2006 Cambridge, MA

## *Translatability*

- 
- Who cares?
  - How translation works
  - Three dimensions of Translatability
    - Content
    - Form
    - Processes

## Who cares?

- Businesses who want to:
  - Increase customer satisfaction
  - Increase sales
  - Increase documentation quality
  - Increase brand consistency
- Decrease stress, headaches, attrition, extra hours
- Decrease costs
- Decrease time to market
- Decrease exposure to liability
- Cope with acquisitions/mergers/outourcing

Tutorial: *Translatable Content* AMTA 2006 ©2006 [inika@nikedillinger.com](mailto:inika@nikedillinger.com)

## Who cares?

- The ROI story
  - [ROI calculator]

Tutorial: *Translatable Content* AMTA 2006 ©2006 [inika@nikedillinger.com](mailto:inika@nikedillinger.com)

# How translation works

- ❶ Documentation Department sends material for translation
- ❷ Localization Department filters out incoming sentences that have already been translated
- ❸ Translation Vendor produces draft translations
- ❹ Translators revise draft translations
- ❺ Documentation department assembles and distributes translations
  
- This is part of the *Content Supply Chain*
  - See: <http://contentsupplychain.blogspot.com>

Tutorial: *Translatable Content* AATA 2006 ©2006 [mike@mikedilinger.com](mailto:mike@mikedilinger.com)

# How translation works

Send material out

- ❶ Documentation Department sends material for translation [e.g., User Manual]
  - What size chunks of material? Paragraphs? Chapters? Books?
  - What else does the translation package include?
  - When is the translation due?
  - Does Marketing use the same process?

Tutorial: *Translatable Content* AATA 2006 ©2006 [mike@mikedilinger.com](mailto:mike@mikedilinger.com)

# How translation works

Filter out already-translated sentences

## ② Localization Department filters out incoming sentences that have already been translated

- Translation memory tools
  - 100% matches : )
    - But still not free
  - Fuzzy matches
    - Alternative formulations of the same content
    - Translation memory is very sensitive to any kind of variation

Tutorial: Translatable Content AMTA 2006 ©2006 mike@mikedittinger.com

# How translation works

Filter out already-translated sentences

- Fuzzy matches
  - *This dialog box is available in more than one application.*
  - *This dialog box is available in more than one contract application.*
  - *Click OK to save changes and return to the application tab .*
  - *Click OK to save the record and return to the application tab .*

Tutorial: Translatable Content AMTA 2006 ©2006 mike@mikedittinger.com

# How translation works

Produce draft translations

## 3 Translation Vendor produces draft translations

- Distribute parts of the project to different translators
- By hand or (increasingly) by machine
- a) Translators read and understand the original text in English
- b) Translators draft "the same information" in L2
- c) Translators revise (with an L2 mindset) so that the result is normal L2 text

Tutorial: Translatable Content

AMTA 2006

©2006

mike@mikedillingor.com



# How translation works

Produce draft translations

## ■ a) Translators read and understand the original text in English

- Translators understand, as possible, given:
  - unknown words,
  - unknown sentence structures,
  - unfamiliar topics,
  - etc.
- Think of how bad it can get for interpreters!



### Heads up!

- Even proficient speakers of English as a Second Language read English at about a 7<sup>th</sup> grade level
- User Manuals are often written at an 8<sup>th</sup> grade level
- "Pre-editing" is needed both for human and for machine translation

Tutorial: Translatable Content

AMTA 2006

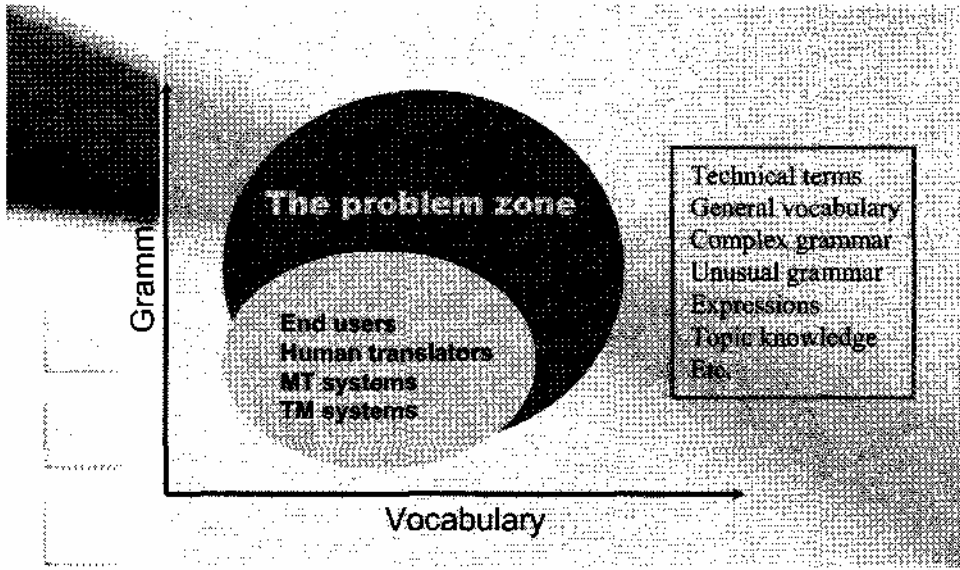
©2006

mike@mikedillingor.com

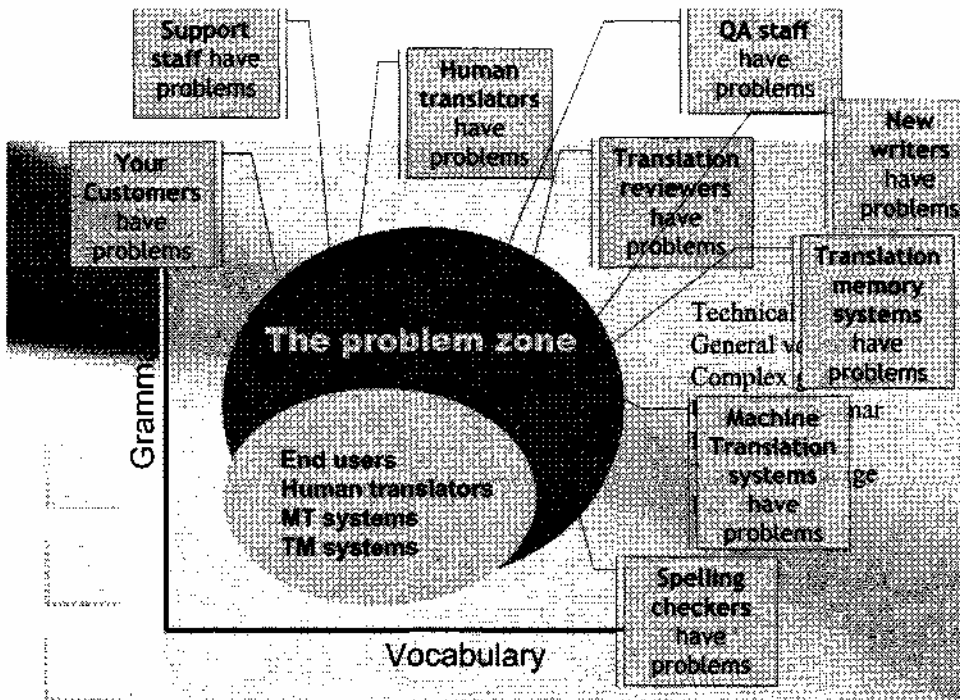


# How translation works

The problem zone



Tutorial: Translatable Content AMTA 2006 ©2006 mikce@makedilling.com



Tutorial: Translatable Content AMTA 2006 ©2006 mikce@makedilling.com

# How translation works


Produce draft translations

- b) Translators draft "the same information" in L2
  - Adjust for in-existent sentence structures
    - "Someone was given a watch."
    - In Mandarin: "A watch was given to someone."
  - Adjust for translations that have more common but irrelevant meanings
    - "The manager has personally passed all the water served here." (In an Acapulco hotel)
    - Teacher > *profesor* or *profesora* (Sp)
    - Eat > *essen* or *fressen* (Ger)

# How translation works

Produce draft translations

- c) Translators revise (with an L2 mindset) so that the result is normal L2 text
  - Systematic use of terminology is difficult for humans
  - Conformity to style guides is difficult for humans

 Heads up!

  - When time and/or training are insufficient, this step suffers most.
  - "Post-editing" is needed both for human and for machine translation



# How translation works

## Produce draft translations

- Some vendors and companies are doing many of the draft translations with Machine Translation technology
  - MT is *much* faster
  - MT is *much* cheaper
  - MT is *much* more consistent
- Machine Translation is both viable and cost effective for draft translations.
  - Quality of the output depends more on writers' conformance to style guides than to limitations of the technology
  - Microsoft is using a writer-direct-to-multilingual-web system for support KB
  - Océ implemented tight integration between writers and translation technologies, cutting localization costs by 60% and translation time by months. Now they've spun the technology off into a new company.
  - SAP, ATT, etc.

# How translation works

## Revise draft translations

### 4 Translators revise draft translations ("post-editing")

- Translation errors
  - Due to lack of linguistic knowledge
  - Due to lack of topic knowledge
  - Due to unclear input
- Familiar words from one language interfere with understanding similar words from the other
  - realize - realizer (Pkg) [accomplish]
  - wish [want] - ??? (Jap)
  - Source-text writers simply can't keep track of these things
- Synonyms
  - Same or different meanings?
  - Have, are made up of, are composed of, contain
- Word lists
  - Guessing the intended meanings is difficult
  - File

# How translation works

More is going on

- **More is going on:**
  - Localization (by hand) of screen shots and figures
  - Localization of software: menus, button labels, error messages, etc.
  - Adjustment of page layout
- **The Localization Department then**
  - Checks that returned translations are complete
  - Samples the translations for quality
  - Discovers that the source text has already changed
  - Organizes files into manuals or help directories
  - Checks functionality of help systems
  - Sends translations out for in-country testing
- **The Documentation Department then**
  - Assembles and distributes the documents

Tutorial: Translatable Content AMTA 2006 ©2006 mike@mikodillinger.com

All too often, you get back just what you sent out:

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?

Tutorial: Translatable Content AMTA 2006 ©2006 mike@mikodillinger.com

Visitors are expected to complain at the office between the hours of 9 and 11 A.M. daily. (In a hotel in Athens)

Please do not feed the animals. If you have any suitable food, give it to the guard on duty. (at a Budapest zoo)

Cooler and Heaters: If you want just condition of warm in your room, please control yourself. (from a Japanese information booklet about using a hotel air conditioner)

To stop the drip, turn cock to right. (In a Finnish washroom)

## Example

Contents of what? What's important?

**Important**  
Read **Contents** Before Using This Product  
Warranty Information

**Important**  
A lire avant d'utiliser ce produit  
Garantie limitée

**Wichtig**  
Bitte vor der Verwendung dieses Produktes lesen  
Beschränkte Gewährleistung:

**Importante**  
Leggere le seguenti informazioni prima  
di usare questo prodotto  
Garanzia limitata

**Importante**  
Lea esto antes de utilizar este producto  
Garancia limitada

**重要**  
製品を使用する前に必ずお読みください  
限定保証

the following information

this

Before using (this product warranty information) or Before using (this product) ??  
+Line breaks are ambiguous!

Tutorial: Translatable Content ANTA 2006 ©2006 mike@mikedlinger.com 20

# Translatability

Three dimensions

## ■ Three dimensions of Translatability

### ■ *Translatable Content*

- How to make the *information* more translatable?

### ■ *Translatable Form*

- How to make the *sentences* more translatable?

### ■ *Process*

- How to deploy processes that facilitate translation?

Tutorial: *Translatable Content*

AMTA 2006

©2006

mike@mikedilinger.com



# Translatable Content

## ■ *Translatable Content is understandable content*

### ■ Control writers' assumptions about readers' background knowledge

#### ■ a) Background knowledge of culture

- Avoid culture-specific language and examples
- Avoid humor, sarcasm, irony, idioms, metaphors
- Avoid political, religious, or local geographic references
- Avoid acronyms - AMTA

- Message therapy or Machine translation?

Tutorial: *Translatable Content*

AMTA 2006

©2006

mike@mikedilinger.com



# Translatable Content

- b) Background knowledge of the product
  - Assume that the reader knows *nothing* about the product.
    - By definition, translators are **NOT** familiar with your product
  - Use specific examples as often as practical
  - Make the text understandable even if the translator can't see the screen
    - Most often, translators don't have a copy of the product
- c) Background knowledge of technology
  - It's extremely easy for people who are familiar with technology to assume that other people are equally familiar with it
  - Tech Support people can very easily supply dozens of examples where this assumption is false
    - Translators are very often not very familiar with technology
  - There's a great collection of unusual assumptions at:
    - <http://inkworks.com/assumg>

## Example

**Customer:** "Hello, is this Tech Support?"

**Tech Support:** "Yes, it is. How may I help you?"

**Customer:** "The cup holder on my PC is broken and I am within my warranty period. How do I go about getting that fixed?"

**Tech Support:** "I'm sorry, but did you say a cup holder?"

**Customer:** "Yes, it's attached to the front of my computer."

**Tech Support:** "Please excuse me if I seem a bit stumped, it's because I am. Did you receive this as part of a promotional, at a trade show? How did you get this cup holder? Does it have any trademark on it?"

**Customer:** "It came with my computer, I don't know anything about a promotion. It just has '4X' on it."

The caller had been using the load drawer of the CDROM drive as a cup holder and snapped it off the drive.

*This is how your documentation sounds to translators and users:*

Is this caused by problems with  
Spelling?  
Punctuation?  
Grammar?  
Terminology?  
Sentence length?  
Formatting?

The writing here follows most style guides.

If the balloons popped, the sound wouldn't be able to carry since everything would be too far away from the correct floor. A closed window would also prevent the sound from carrying, since most buildings tend to be well insulated. Since the whole operation depends upon a steady flow of electricity, a break in the middle of the wire would also cause problems. Of course, he could shout, but the human voice is not loud enough to carry that far. An additional problem is that a string could break. Then there could be no accompaniment to the message. It is clear that the best situation would involve less distance. Then there would be fewer potential problems. With face-to-face contact, the least number of things could go wrong.

## Next steps

### ■ Tools for Translatable Content

- None for monitoring content explicitly yet
- Content type checking in XML
- BUT term and style checking help catch unplanned terms and sentence types

### ■ Techniques

- Training
- Audits
- Know your clients
  - Build empirical (not heuristic) reader profiles
  - FAQ hit rates from support staff and support pages
  - Sales, Marketing, Product Managers
  - Direct feedback and suggestions
  - User conferences
  - Feedback from translation vendors
- Foster community-driven documentation

### ■ Further reading

- Sun Microsystems (2003), *Read me first!*, Chapter 7

# Translatable Form

## ■ Form

### ■ How to make the sentences more translatable?

- Use translatable vocabulary and terminology
- Use translatable sentence structure
- Writing "for MT"?
- Other

# Translatable Form

## Vocabulary and terminology

### ■ Translatable vocabulary and terminology

- Use a word in the same way throughout your documents
  - Same part of speech, same meaning
    - Record, record
    - Your work, to work
    - Especially in the same sentence
  - Do not use synonyms
    - Translators assume that synonyms are different words with different technical senses
- Avoid unusual (for the reader) word usage
  - To select a labor, *select Select Record*
  - Do not invent new words or word senses
  - Be careful with technical terms that are also common everyday words
    - Class, pool, windows, labor, etc.

# Translatable Form

Vocabulary and terminology

- **Translatable Vocabulary and terminology (2)**
  - Avoid colloquial and regional expressions
    - Especially two- and three-word verbs
      - *If you get hung up on the details, ...*
  - Avoid or document word ambiguity
    - Bundle glossaries with texts for translation
    - Bundle preferred translations of terms with texts for translation
  - Provide example sentences for word lists
    - File as in...
  - Manage your terminology
    - Technical and non-technical words, as well
    - Know which terms are most frequent
      - These are the ones that need close revision
      - Reduce the number of words used only once or twice

Tutorial: Translatable Content AATA 2006 ©2006 mlke@mlkediting.com 29

## Next steps

- **Tools for Terminology Management**
  - MultiTerm
  - WordSmith
  - Spreadsheet on the intranet, at least
  - Integrated term and style checking
    - acrocheck, CLAT, Boeing / HyperSTC
- **Techniques**
  - Training
  - Terminology committee
  - Audit: key-terms analysis
- **Further reading**
  - Wright, SE & Budin, G. 1997. *Handbook of Terminology Management* (2 vols). Amsterdam: John Benjamins.

Tutorial: Translatable Content AATA 2006 ©2006 mlke@mlkediting.com 30



# Translatable Form

## Sentence structure

### ■ Translatable Sentence structure

#### ■ "Readability plus"

#### ■ Avoid unusual or complex sentence structures

- Aim for 7<sup>th</sup>- or 8<sup>th</sup>-grade reading level
- Much current documentation is at an 11<sup>th</sup>-grade level
- As much for translators as for translation memory systems

#### ■ Adverbs

- Avoid directional adverbs
  - *Drag your cursor across the screen.*
    - *>Make your cursor cross the screen by dragging it.*
  - *Override, underachieve*

# Translatable Form

## Sentence structure

### ■ Translatable Sentence structure

#### ■ "Readability plus"

#### ■ Prepositions

- Use prepositions only for spatial and temporal meanings
- 90% of all prepositions:  
*of, in, to, for, on, with, at, by, from, as, into, about, after*
- Put prepositional phrases next to the item that they complement
  - *Remove the filler panel from the slot with the pliers*
  - *Use pliers to remove the filler panel from the slot*

# Translatable Form

## Sentence structure

### ■ Translatable Sentence structure

#### ■ "Readability plus"

#### ■ Conjunctions

- Take extra care with *and* and *or*
  - Parallelism is essential
- Sorry, only one conjunction per sentence

#### ■ Verbs

- Do not omit "that" for verb complements
- Do not use passive sentences
- Do not use two- or three-word verbs
- Do not change commonly used transitivity
  - Based on the current item, one of the following applications launches: if the item is a valve, the Valves application launches.

# Translatable Form

## Sentence structure

### ■ Translatable Sentence structure

#### ■ "Readability plus"

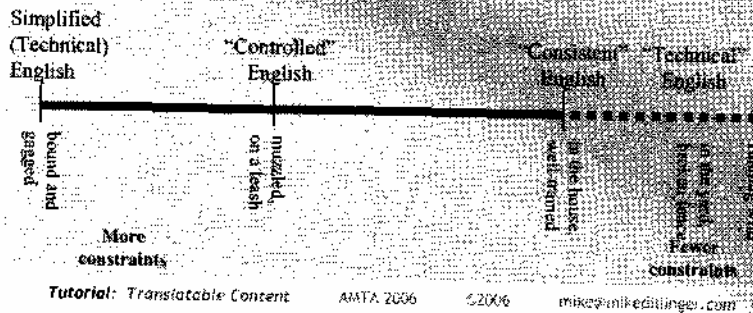
#### ■ Nouns

- Do not use noun stacks with more than 2 elements
  - Each route stop may have its own job plan, which the system copies to the child route stop work order
- Avoid -ing nouns
  - [Drinking water] is good for you.
  - Cooking rocks! (Rebecca Ray cookbook)
    - -how to cook rocks
    - -rocks for cooking
    - -cooking is cool

# Translatable Form

Writing "for MT"?

- Writing "for" machine translation?
  - The myth: MT requires special writing
  - The reality: People require special writing
  - Reuther (2003)
    - The rules are pretty much the same for human translators and for machine translation systems (and for end users)



# Translatable Form

Writing "for MT"?

- Seven 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
- 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

## Next steps

- **Tools for Translatability Management**
  - *Integrated* term and style checking
    - acrocheck, CLAT, Boeing / HyperSTE
- **Techniques**
  - Training: authors don't know about this
  - Training: improve style guides
  - Audit: Translatability analysis
  - Audit: Reuse analysts
  - Planning: identify and prioritize issues
- **Further reading**
  - Kohl, J. (1999). "Improving translatability and readability with syntactic cues" Technical Communication, vol. 46, no. 2, pp. 149-166

## Translatable Form

*Other topics*

- **Lay out the text to permit different text lengths, different paper sizes**
  - To avoid DTP in the target language
- **Be careful with text variables**
  - In some contexts, they have to be capitalized or not
- **Programmers HAVE TO put all strings in resource files, not in the code**
  - Who revises *their* English?
  - Improve Development Environments for better string management

# Translatable Processes

## ■ Processes

- How to make processes that facilitate translation?
- a) Reduce the amount of text for translation
- b) Make translation itself easier
- c) Improve the processes that support translation

# Translatable Processes

*Reduce the amount of text for translation*

## ■ Reduce the amount of text for translation

### ■ Write less

- Write only what the clients need
  - Reassess relevance
- Content on demand?
  - Content inventory vs. JIT

### ■ Reuse more

- Cross-departmental reuse
- Topic reuse
- Sentence reuse
- "Single-sourcing"
- Simplify updates

## Next steps

### ■ Tools for Content Reduction

- Single-sourcing
  - Ex: AuthorIT, Documentum, Idiom

### ■ Techniques

- Use Machine Translation
- Planning: Develop a reuse map
- Training: Do a content lifecycle analysis
- Training: Writing for reuse
- Audit: Reuse analysis
- Know your clients

### ■ Further reading

- Ament, K. (2003). *Single Sourcing: Building modular documentation*. Norwich, NY: William Andrew.
- Rockley, A. (2002). *Managing Enterprise Content: A unified content strategy*. New York: New Riders.

Tutorial: *Translatable Content* AMTA 2006 ©2006 mike@mikedillinger.com

## Translatable Processes

*Make translation itself easier*

### ■ Make translation itself easier

- Write for translatability
  - Mine translator feedback
  - Wrest control of interface strings from engineering
- Include multilingual glossaries
  - Part of the translation package
- Standardize tools and file formats
- Keep translation memories up to date
  - Maintain hit-rate data
- Anticipate translation problems
  - Cut down on screen shots
  - Lay out for translations that are 30% longer
  - Make updates easier

Tutorial: *Translatable Content* AMTA 2006 ©2006 mike@mikedillinger.com

## Next steps

- **Tools for Translatability Management**
  - *Integrated* term and style checking
    - acrocheck, CLAT, Boeing / HyperSTE
- **Techniques**
  - Training: authors don't know
  - Training: improve style guides
  - Audit: Translatability analysis
  - Audit: Reuse analysis
- **Further reading**
  - Kohl, J. (1999). "Improving translatability and readability with syntactic cues" *Technical Communication*, vol. 46, no. 2, pp. 149-166

Tutorial: *Translatable Content* AMTA 2006 ©2006 mike@mikedilling.com

## Translatable Processes

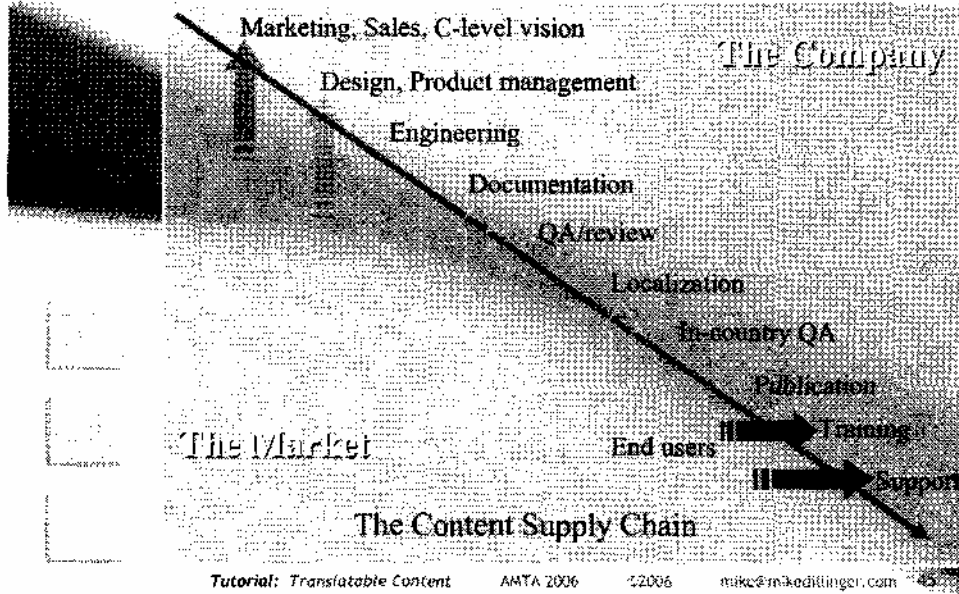
*Improve the processes that support translation*

- **Improve the processes that support translation**
  - **Adopt a supply-chain mentality!**
    - Communicate & Coordinate
    - Know your suppliers upstream
    - Know your clients downstream
    - Anticipate their needs systematically
      - Success means things are easier for everyone
  - **Automate the Content Supply Chain**
    - Workflow software is essential
  - **Simplify the Content Supply Chain**
    - <insert controversial discussion here>

Tutorial: *Translatable Content* AMTA 2006 ©2006 mike@mikedilling.com

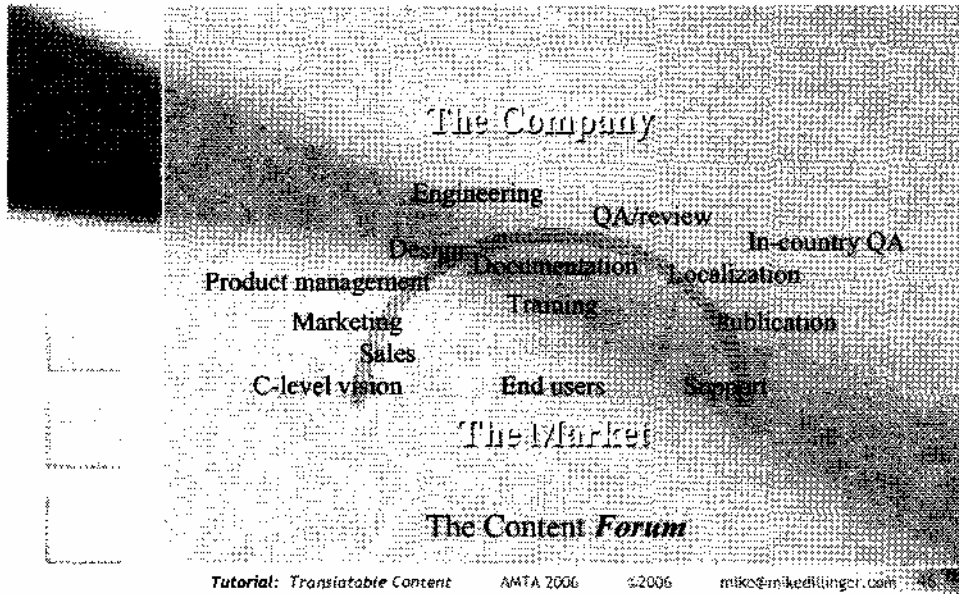
# Translatable Processes

Improve the processes that support translation



# Translatable Processes

Improve the processes that support translation





## Next steps

- Tools for optimizing processes
  - Content Mgmt Systems, Globalization Mgmt Solutions, Enterprise Content Mgmt, etc.
  - Wikis, a blackboard approach
- Techniques
  - C-level involvement
  - Six Sigma, ISO 9000, etc.
  - Access engineers' development environment
  - Access support's data on FAQs
- Further reading
  - Friedman, T. (2006). *The World is Flat*. New York: Farrar, Straus & Giroux.

Tutorial: *Translatable Content*

AMTA 2006

©2006

mike@mikedillinger.com



## Summing up

- Who cares about translatability?
- How translation works
- Three dimensions of Translatability
  - *Translatable Content*
    - How to make the *information* more translatable?
  - *Translatable Form*
    - How to make the *sentences* more translatable?
  - *Process*
    - How to deploy processes that facilitate translation?

Tutorial: *Translatable Content*

AMTA 2006

©2006

mike@mikedillinger.com



# Thanks

- Thanks for your attention
- More questions?