



Sixth Biennial Conference of the  
Association for Machine Translation in the Americas

## Tutorial Notes

# Mastering Machine Translation (MT) output

Jeff ALLEN  
MultiLingual Computing & Technology magazine  
Member of Editorial Advisory Board

*September 28, 2004  
Georgetown University  
Washington, DC*

MultiLingual



## Outline

- ◊ Tutorial Objectives
- ◊ Human Translation production cycle (TEP cycle)
- ◊ Translation approaches (effort vs. expected result)
- ◊ Translation system types
- ◊ Mapping commercial MT tool features to translation approaches and needs
- ◊ MT production cycle
- ◊ How to think like an MT system
- ◊ MT Dictionary building principles
- ◊ MT postediting with commercial MT software

MultiLingual



## Tutorial Objectives

- Briefly describe the different MT engine types, yet focusing on the RBMT model as the most common commercial MT type.
- Describe the different types of translation approaches
- Describe the human translation production cycle
- Describe a non-interactive MT cycle (pure content gisting)
- Explain the human-interactive MT production cycle
- Identify which translation features/modules in various MT products correspond to different types of translation approaches and needs
- Demonstrate how to use the different translation features / modules in a couple of products
- Clearly prove that poor quality translation output is not limited to system ability, but really depends on user knowledge of translator-focused features
- Describe how commercial MT engines process (parse) the data
- Show participants how to master the principles of basic vs complex dictionary entries
- Demonstrate how to use postediting features built-in to MT software



## Specific Points to note

- This tutorial concerns Machine Translation (MT) (also called Automated or Automatic Translation) software and systems.
- This tutorial does not cover simple bilingual dictionary systems. It is full-fledged MT software.
- I “intentionally” refer to Translation Memory (TM) tools as the commercial variants of Example-Based MT (EBMT) system. From a pure MT developer’s point of view, this is not 100% accurate, yet from a user’s perspective, TM tools are the only existing commercial version of EBMT-like systems available today.





## Factors to consider for using MT system/software

This is a list of guideline topics to evaluate one's real need

- ▣ Translation approach
- ▣ Documentation (or speech) type (procedures, descriptive, diagnostic troubleshooting, etc)
- ▣ Domain/field (medical, automotive, insurance, financial, etc)
- ▣ General text or very technical text
- ▣ User/audience type
- ▣ Number and types of language directions
- ▣ (Semi-)Automated translation system types
- ▣ Levels of human-machine interaction
- ▣ Means for capturing human knowledge and recycling it
- ▣ Mapping system features/modules to translation approach types to be used



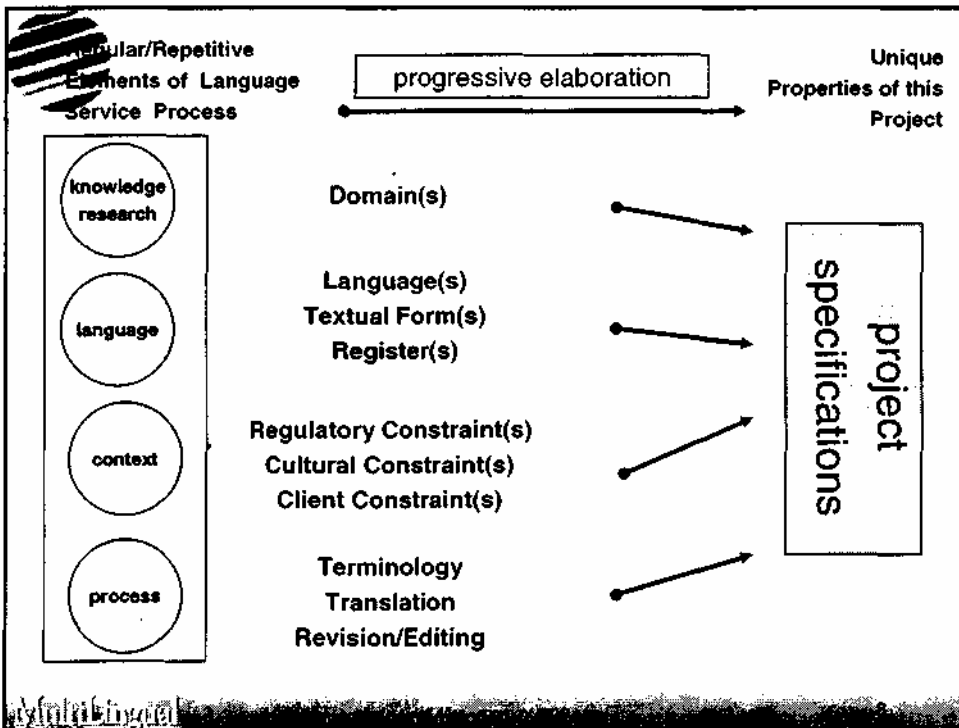
## Some needs analysis questions

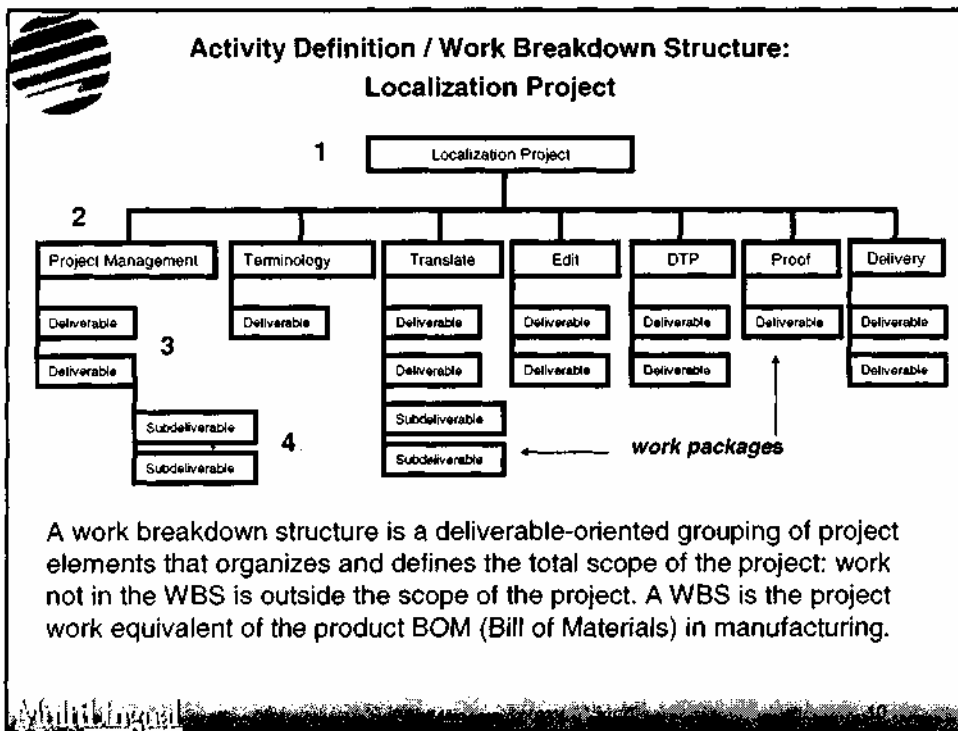
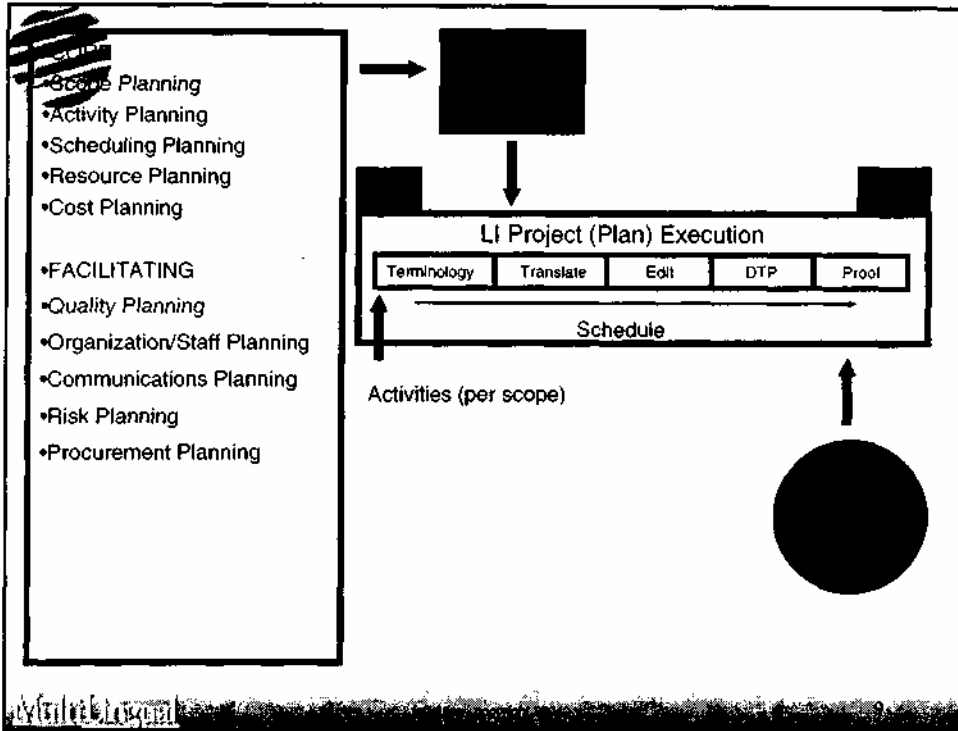
- ▣ What translation approach(es) to use (Inbound, Outbound, or both)?
- ▣ What types of documentation (or speech) to translate?
- ▣ What is the expected throughput quality?
- ▣ What is the threshold of error tolerance by users?
- ▣ What is the level of language (ie, language register) to be used?
- ▣ What are language skills of the MT users?
- ▣ What are the types of (semi-)automated translation systems? Do these fully or partially correspond to the need?
- ▣ Will human-machine interaction be part of the intended process?  
Can the existing software/systems cover the necessary interaction?
- ▣ What human knowledge capture/insertion channels can be used?
- ▣ What features and modules in current commercial translation software (MT, TM, etc) correspond to the translation needs, and which features/modules do NOT correspond?



## The TEP Cycle

- What processes have human translators used for years for their translation workflow cycle? (the Translate/Edit/Proof cycle)
- Stages to the overall Translation process (including TEP)
  - Selection of Translator(s)
  - Delivery of original texts to translator(s)
  - Test
  - Cultural Analysis
  - Translate, Edit, Proof
  - HTML Coding & Review
  - Delivery of the translation
  - Uploading
  - Review and Feedback
- Some references for TEP process
  - <http://appling.kent.edu/ResourcePages/Courseware/Localize/Week14/Localization-Week-14.ppt>
  - <http://www.stcwestcoast.ca/conf/proceedings/slides.pdf>
  - <http://www.1st-translation.com/general.html>
  - [http://www.stc.org/50thConf/Session\\_Materials/dataShow.asp?ID=252](http://www.stc.org/50thConf/Session_Materials/dataShow.asp?ID=252)
  - [http://www.translate.com/technology/multilingual\\_standard/quality\\_in\\_localization.html](http://www.translate.com/technology/multilingual_standard/quality_in_localization.html)
  - [http://www.stc.org/50thConf/Session\\_Materials/dataShow.asp?ID=251](http://www.stc.org/50thConf/Session_Materials/dataShow.asp?ID=251)
- The next three slides are taken (by written permission) from a tutorial course by Gregory Shreve, Kent State University
  - <http://appling.kent.edu/ResourcePages/Courseware/Localize/Week14/Localization-Week-14.ppt>







## MT software affected the cycle

- Translation software applications (MT as well as TM) started being implemented in the 1980s, and this affected the traditional TEP cycle.
- There were promises in the 1980s and 1990s to provide 100% MT with no postediting
  - ALLEN, Jeffrey. March 2004. Thinking about machine translation; several questions to ask yourself when you read an article about MT technologies. In special supplement of Multilingual Computing and Technology magazine, Number 62, March 2004. (<http://www.multilingual.com/machineTranslation62.htm>) Refers to statements made in article in Multilingual Computing & Technology, Number 5, Vol 5, Issue 2.
- Arrival of online MT systems (starting with AltaVista Babelfish) began a new era of Information/Content gisting translation
  - ALLEN, Jeffrey. 2000. The Value of Internet Translation Portals. In International Journal for Language and Documentation (IJLD), Issue 6, August/September 2000, pp. 45-46. (article available online at <http://www.geocities.com/mipostediting>)
- This began the new focus on distinguishing between *Inbound* (Assimilation/Gisting) vs. *Outbound* (Dissemination/Translation for Publication) Translation needs and processes.

www.multilingual.com



## Translation approaches

### Inbound Translation approach

- MT Browsing-Information/Content Gisting (no post-editing)
- Note: I refer to this as "Push-button Inbound MT"
  - ▶ online (published and frozen) documents (general: newspapers, brochures; customer support docs and e-mail, white papers)
  - ▶ e-mails
  - ▶ internet / intranet web pages
  - ▶ MS Powerpoint presentation for understanding only
  - ▶ draft translation of doc to determine relevance and make preliminary decision
  - ▶ restaurant menus on trips
  - ▶ way finding directions received in foreign language
  - ▶ voice-driven online help menus (phone voice box, immigration service, health dept, hospital, pharmacy, etc)

www.multilingual.com



## Translation approaches

### Inbound Translation approach

- Rapid Postediting
  - ▶ European Commission (EC) Post-édition rapide (PER) / Rapid PE (RPE) Translation service to translate urgent texts "for information only"
    - *internal meeting working papers*
    - *meeting minutes*
    - *technical reports or annexes*
    - *short-life documents*
    - *etc*



## Translation approaches

### Outbound Translation approach - Translation for publication

- MT with no post-editing (PE)
  - ▶ Tourism: basic survival communication (travel and lodging reservations, interactive way finding, purchase food necessities at bakery, grocery store, etc). A PocketPC or Palm-based MT software application is appropriate hardware/software configuration for this type of usage.
  - ▶ E-mail: forwarding e-mail messages to non-speakers of the source languages of the source text.
  - ▶ Customer support : rapid internal company e-mail queries from set of communication guidelines + training



## Translation approaches

### Outbound Translation approach - Translation for publication

- Partial PE (minimal and/or rapid)
  - Urgent huge document translation requests impossible to complete without translation tools (remember the 1998 Clinton-Starr report in 1998: approx 500 pages).
    - *Real example cited by a professional translator in 2004: translation of an 80-page user manual in less than 8 hours can only be done by computer-aided translation.*
  - User/Service manuals: Caterpillar Operation manuals and Service manuals, General Motors user manuals, Microsoft user documentation, Sun Microsystems user documentation, etc
  - Draft translation of document as preliminary step to determine translation need before seeking high-quality human translation:
    - *Real example in 1997: a book editor asked me if MT systems could be used to scan review the content of incoming books in foreign languages. This would be first step to determine if content is generally valid. Then proceed with human translation of entire book afterward (most likely not even using the MT'd text).*
  - Customer support: need to send immediate reply by e-mail based on source-only documentation. Must have "technical support" knowledge, fully bilingual skills + basic mastery of translation tool
    - *Real example in 2003: customer need to rapidly produce multilingual versions of software workaround explanations based on existing technical documentation and release notes in at least one language.*



## Translation approaches

### Outbound Translation approach - Translation for publication

- Full Postediting (PE): examples of where I have used it as the process to translate materials
  - Marketing brochures (ALLEN, Jeffrey, 2001. Postediting: an integrated part of a translation software program. In Language International magazine, April 2001, Vol. 13, No. 2, pp. 26-29. <http://www.geocities.com/mtpostediting/Allen-LI-article-Reverso.pdf>)
  - User documentation (Reverso Mac Software User manual, Softissimo, 2001)
  - Articles (ALLEN, Jeff, 2002. The Bible as a Resource for Translation Software: A proposal for Machine Translation (MT) development using an untapped language resource database. In Multilingual Computing and Technology magazine, Number 51, Vol. 13, Issue 7, October/November 2002. Pp. 40-45. English original version available at <http://www.multilingual.com/allen51.htm> & MT + Full Postedited French translation at <http://www.editionsclie.com/bcl/presse/article1/allen-mltc51-fr.htm>)
  - Response for Information (Mycom International, March 2004, described in User Focus paper to be presented at AMTA-2004 by Jeff Allen)
  - Software Acceptance Test Plan (Mycom International, March-April 2004, described in User Focus paper to be presented at AMTA-2004 by Jeff Allen)





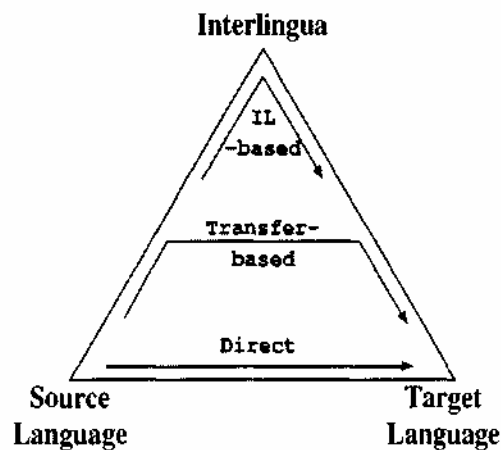
## Full Postediting (PE) with commercial MT Software programs

- The rest of this tutorial focuses on using appropriate MT software for your translation approach(es)
- Do NOT use free online MT portals to perform full PE tasks. If you do, it shows 1 of 2 things:
  - 1) you are completely ignorant of the inappropriate use of your energy (so you have no right to complain about the MT output quality) and/or
  - 2) you are too cheap to invest in something that will be productive whereas there are ranges of products with different prices that correspond to different needs (so again you have no right to complain)
- Choose your MT tool(s) (incl. brand, version, and modules) carefully based on current needs but also consider possible future needs that might currently be unexpressed or undefined



## Machine Translation System types

- Interlingua-based
- Transfer-based
- Direct
  - Example-based
  - Statistical





## Machine Translation system types

- **RBMT (Rule-based MT):** majority of commercial MT systems (SYSTRAN, PROMT, Reverso, SDL Transcend, Babylon, etc)
- **EBMT (Example-based MT):** Translation Memory (TM) applications are EBMT-like. Several academic and government-funded EBMT research systems. Also Microsoft's Data-Driven MT.
- **KBMT (Knowledge-based MT):** thematic/semantic role/category approach in some customized industrial MT systems (CMU KANT) & few commercial MT systems (CIMOS, Translingo)
- **S(B)MT (Statistical-based MT):** many academic and government funded research systems, very few commercial MT systems (Language Weaver)
- **MEMT (Multi-engine MT):** government funded research systems & some customized industrial MT systems.
- **Hybrid Translation software:** Many commercial MT & TM vendors now including TM compatible modules in MT software and MT-like features in some TM software.



## Overview of modules / features of some commercial MT systems

Which features / modules in MT systems correspond to various translation approaches / needs?

### Points to note:

- All commercial MT software vendors provide a range of MT products which contain different types of translation modules.
- Most commercial MT software products contain one or more modules which correspond to different types of translation approaches and needs.
- Although a translation user might use a module of an MT translation application for one need, s/he might use a different module for other needs (due to different needs/tasks).



## Overview of modules / features - MT systems

- Reverso Expert / Pro:
  - Inbound: Webview, Reverso Express, QTrans, MS Plug-ins, File Translator, Reverso Perso, Reverso Corporate Intranet
  - Outbound (basic): Reverso Express, QTrans, MS Plug-ins, File Translator, Reverso Perso, Dictionary Editor (basic level)
  - Outbound (advanced): Promt, Dictionary Editor

See my review of the product at  
<http://www.geocities.com/jeffallenpubs/swreviews.htm>



## Overview of modules / features - MT systems

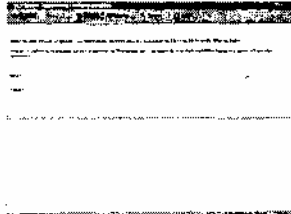
- ePROMT Expert / Professional (PROMT XT)
  - Inbound: PromtE Web Translator, Clipboard translator, MS Plug-ins, File Translator, PROMT Corporate Intranet, PROMT Standard, PDF Plug-in
  - Outbound (basic): PROMT Standard, QTrans, Dictionary Editor, File Translator
  - Outbound (advanced): PROMT, Dictionary Editor, PROMT for Trados, Associated Memory

See my review of the product at  
<http://www.geocities.com/jeffallenpubs/swreviews.htm>

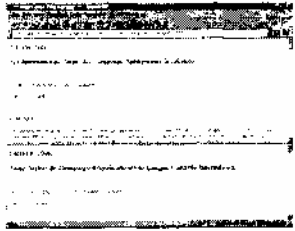


## Visual examples of MT software modules – PROMT & Reverso

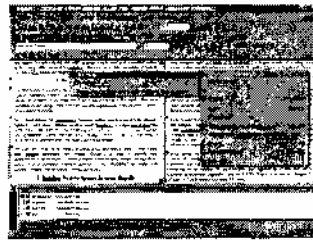
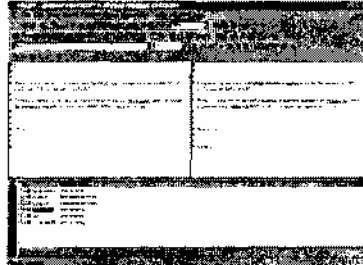
Clipboard translator



Web translator

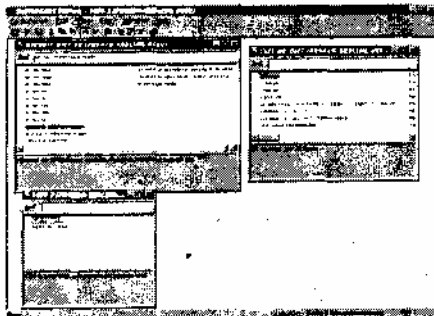


PROMT interface

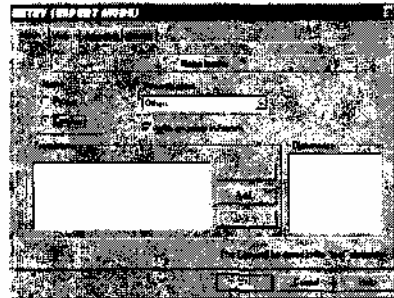


## Visual examples of MT software modules – PROMT & Reverso

Dictionary editor



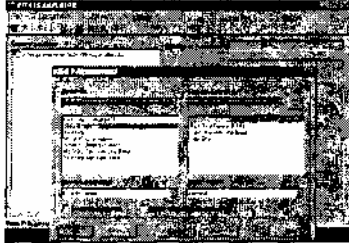
Dictionary entry dialog



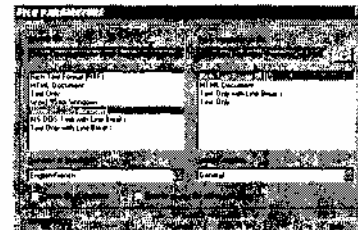
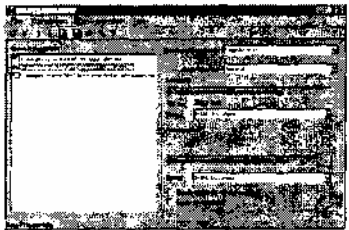
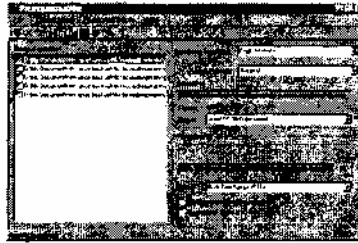


## Visual examples of MT software modules – PROMT & Reverso

File Translator



File Translator

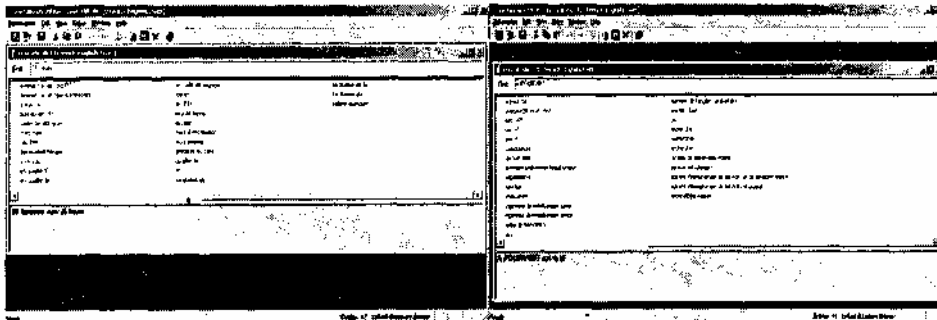


Antitext.com



## Visual examples of MT software modules – PROMT & Reverso

### User Dictionary entries



Antitext.com



## Overview of modules / features - MT systems

- SYSTRAN v4 Premium
  - Inbound: SYSTRAN Toolbar (STB) MS plug-ins, SYSTRAN Clipboard Taskbar (SCT), SYSTRAN Multitranslate Utility (SMTU)
  - Outbound (basic): SYSTRAN standard, Clipboard Taskbar (SCT), Multitranslate Utility (SMTU)
  - Outbound (advanced): Translation Professional Manager (STPM), Dictionary Manager (SDM)

See my review of the product at

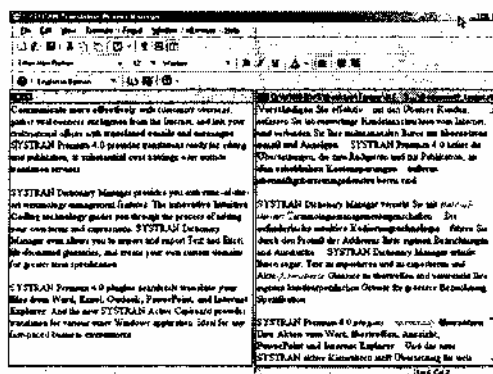
<http://www.geocities.com/jeffallenpubs/swreviews.htm>

- SYSTRAN v5 Professional Premium: I am currently reviewing it



## Visual examples of MT software modules – SYSTRAN Premium v4

### User Dictionary entries







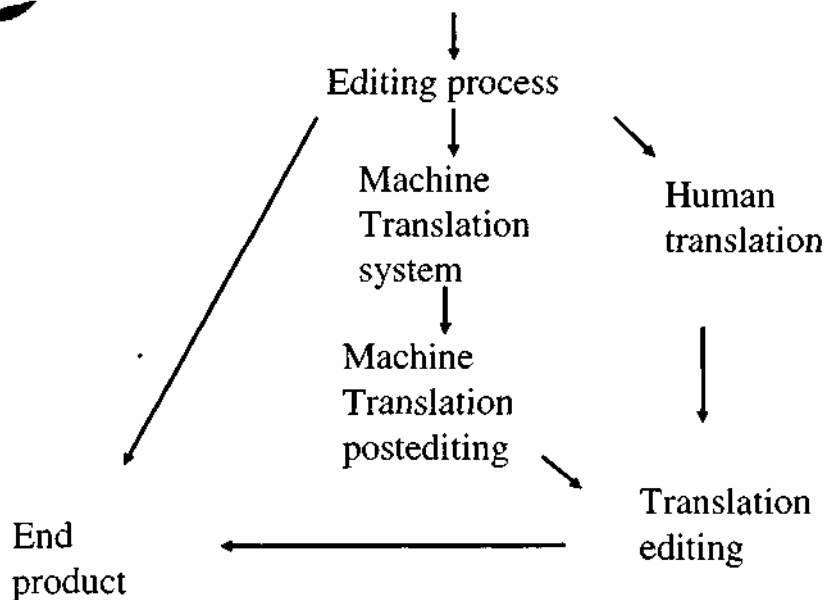
## Examples of Previous MT integration cycles

- Now that we have had an overview of 1) the different types of translation approaches, 2) the different types of MT systems, and 3) how the different commercial MT software modules correspond to the first two points, the big question is how to integrate an Outbound-focused MT software program into an environment with human translators
- The following slides provide some examples of examples of how MT has been implemented into translation workflows.
- The first two slides are from Jeff Allen (1999, Implementing Controlled Language and Machine Translation in the Automotive Industry, Multilingual TOPTEC symposium, Amsterdam, Society for Automotive Engineering).
- These slides show indicate the lack of considering MT postediting as the translation stage of the traditional TEP process and not maintaining the full TEP cycle.

cyberlingua

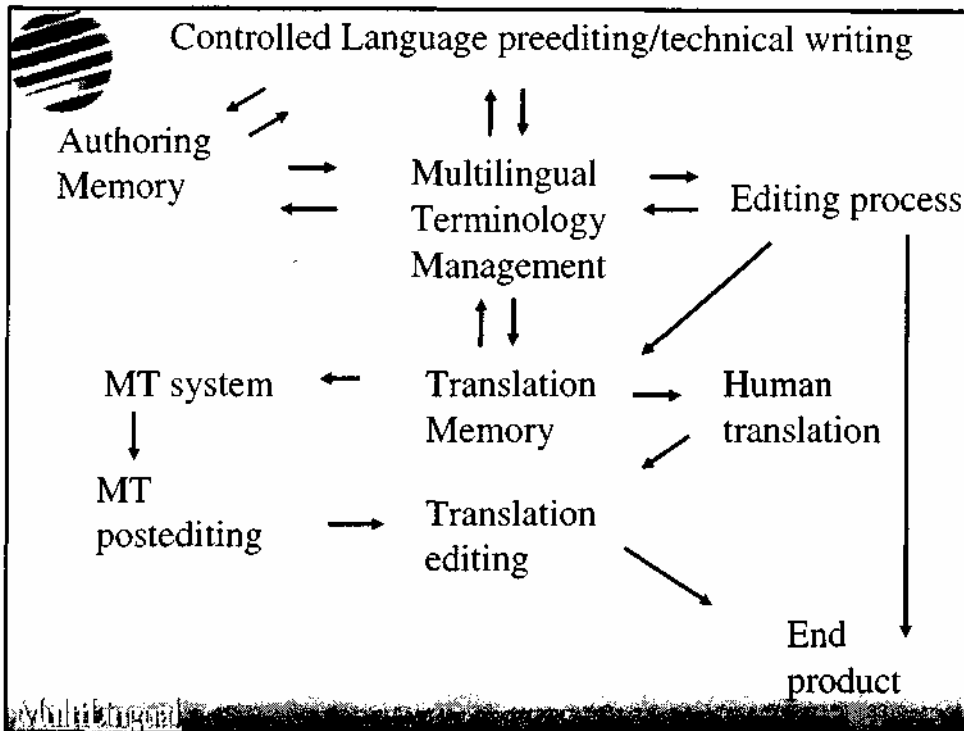


## Controlled Language preediting/technical writing



cyberlingua





- Psychological-Cognitive Workflow task approach MT Technology Integration**
- Most MT implementations in the past have been simply technology focused. They have not usually taken into consideration the user's real-life workflow and set of tasks, and the interactive processing that takes place.
  - After having developed and implemented several different MT systems, and having trained hundreds of users of them, I have modified my approach to how to implement and train users on such software.
  - My new approach is more task-focused with the purpose of highlighting the cognitive tasks that the translator must conduct.
  - Therefore, I present some diagrams about integrating MT systems (or any technology) into the workplace.
  - The next set of slides are adapted from Jeff Allen (1999, Implementing Controlled Language and Machine Translation in the Automotive Industry, Multilingual TOPTEC symposium, Amsterdam, Society for Automotive Engineering).



## Technology

T		V
e		e
c		r
h		t
n		i
o		c
l		a
o		l
g		
y		

www.kit.edu



## Communication - Relationships

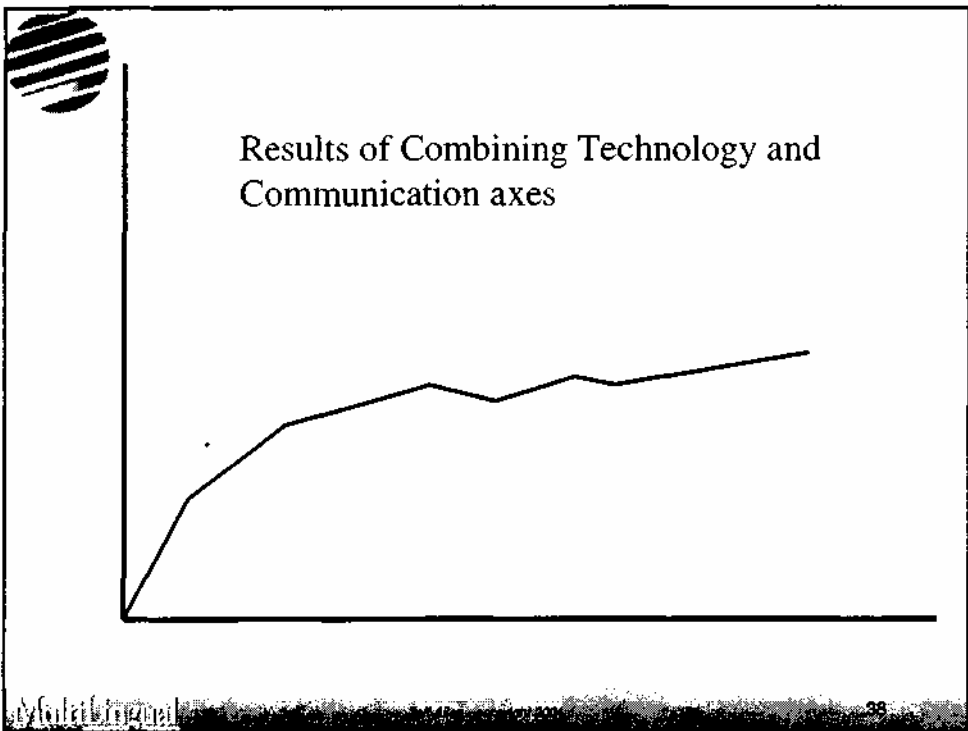
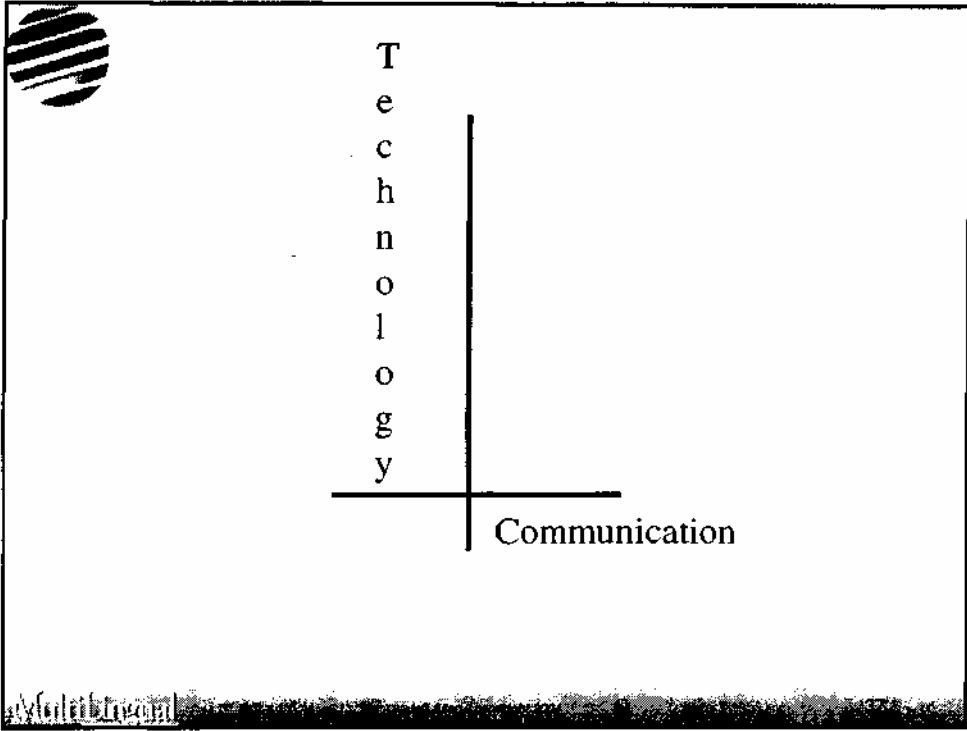
---

### Communication Relationships - Horizontal

3 Types of Communication which are in terms of relationships:

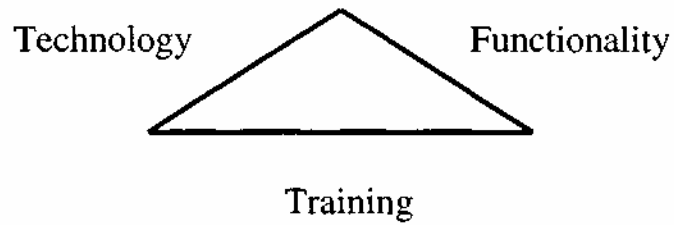
- Human-Human Communication
- Computer-Computer Communication
- Human-Computer Interaction (HCI)

www.kit.edu





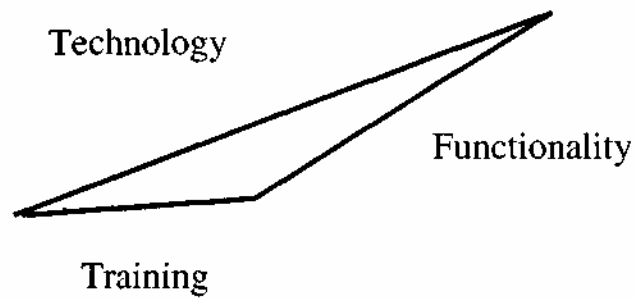
## Technology Integration Triangle (Allen, February 1998)



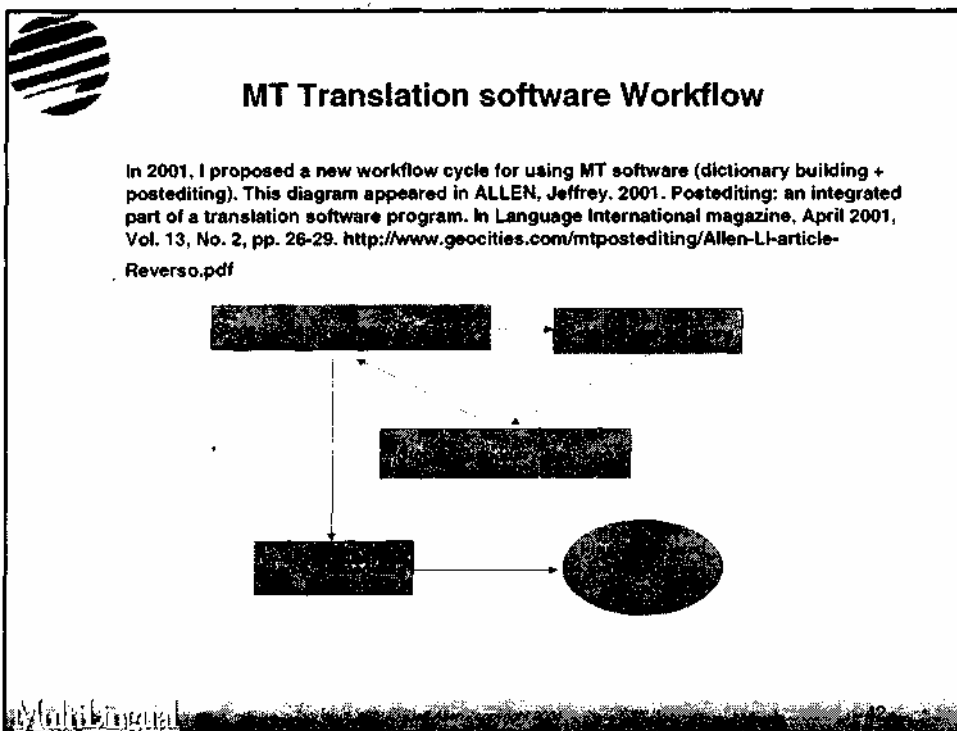
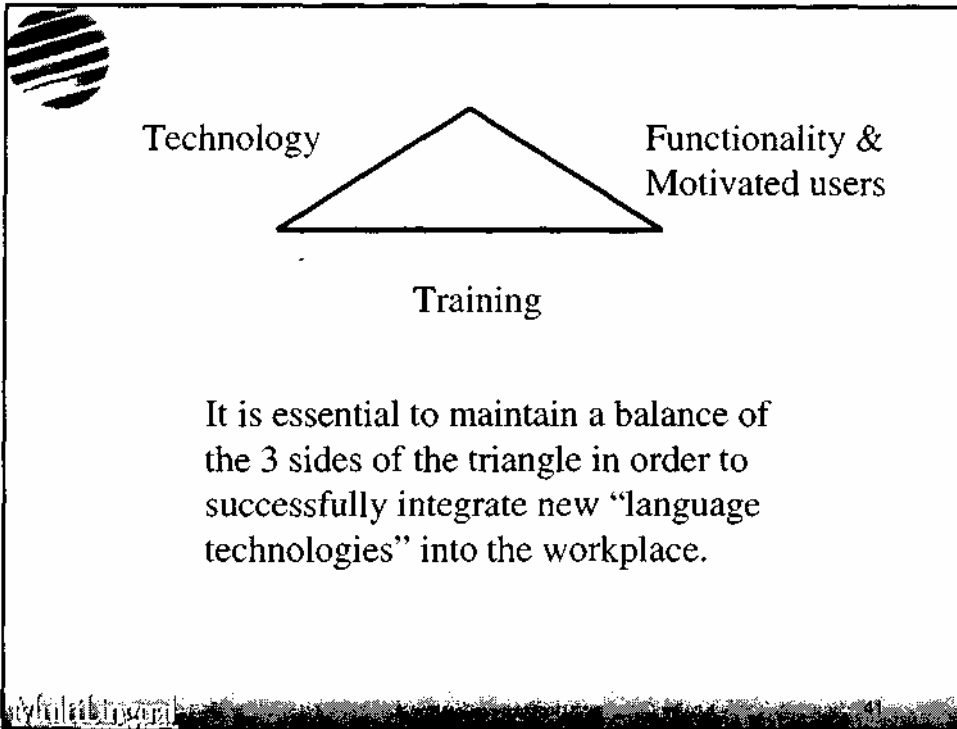
www.ictjournal.com



## Common case scenario



www.ictjournal.com





## Thinking like an MT system

- You need to think with the Verb as the center.
- You need to think in terms of Noun groups (technical terms) being processed first. If Noun group / technical terms are identified and recognized by the system, then they prevail in the processing (parsing) of the sentence.
- Once the Noun Groups are processed, then the individual words are processed.
- You need to think linearly. The elements at the beginning of the sentence get considered before the elements toward the end of the sentence.



## MT Dictionary Types

- **General dictionaries**
  - A general dictionary is provided for each language translation direction.
  - It is the default dictionary (ie, the last dictionary consulted by the system). If a term is in no other dictionary type, and not in this dictionary, then the term is considered to be an Unknown word/term.
  - These entries are View-only (Read-only permission).
- **Pre-packaged/Bundled Specialized/Customized dictionaries (per domain/field)**
  - These dictionaries are created and distributed/sold by the MT vendor company, or by their partner institutions.
  - These can be general specialized dictionaries (for a field in general) or customized dictionaries for a specific customer.
  - These dictionaries override the general dictionary entries.
  - These entries are View-only (Read-only permission).
- **User dictionaries**
  - These dictionaries are created by the user.
  - The user can create as many user dictionaries as wanted.
  - These dictionaries can usually be set in a specific order of priority with the dictionary manager of the MT software.
  - These dictionaries override the general dictionary entries AND the specialized dictionary entries.
  - These entries have Read and Write permissions for the user.



## MT Dictionary building tips

- **Rapidly create multiple related entries**
- **Double-check the dictionary class**
- **Set the order priority of your dictionary entries.**
- **punctuation and special symbols**
- **back up your dictionary files on regular basis**
- **capital letters**
- **spelling variants**
- **Singulars and plurals**
- **multiple variants**
- **Accented characters in the dictionary**



## Dictionary Exercises



## Postediting Features

- Paragraph alignment & side-by-side source and target text viewing
- Translation alternatives / variants
- Translation algorithms / pragmatic customized variables
- Self-learning of multiple translations
- Colored underlining of words or phrases
- Highlighted sentence segments for quick analysis
- List of unknown words
- Not-to-translate words or phrases
- Printing in bilingual format
- Word Mover



## Postediting Examples