Building a Business Case for MT

11-731 Machine Translation

What Makes a Useful MT System?

- Maximum accuracy?
- Maximum # languages?
- Most general domain?
- Lowest development cost?
- Lowest update cost?
- Answer: ‘usefulness’ can only be defined relative to specific requirements of a given domain

Judging Usefulness of MT

- Generality in coverage vs. accuracy of output
- Available resources vs. required development effort, schedule
- Human interaction at run-time
- Levels of output acceptability

In General...

- What makes a good language system?
  - high accuracy / minimal ambiguity
  - "right" level of coverage
  - extensible architecture
  - good support tools

In General...[2]

- What makes a good software system?
  - good documentation
  - well-designed interface
  - robustness
  - modular, extensible knowledge
  - maintainability / support tools
  - well-defined / well-supported workflow
Customer Requirements

- Translation volume? (pages/year)
- Translation quality? (rough vs. publication quality)
- Development time?
- System update? (volume, freq)
- Operational cost? (HW/SW)
- Human (postediting) cost?
- Expected return on investment?

Example: KBMT (KANT)

- Volume: high (1000’s pages/year)
- Translation quality: publication
- Development time: 1-2 years
- System update: varies per domain
- Operating Cost: varies per update
- Human cost: pre-/post-editing
- Expected return on investment: aggressive

Example: Rapid Deployment (DIPLOMAT)

- Translation volume: very low
- Translation quality: rough
- Development time: 1-2 months
- System update: constant
- Operational cost: low
- Human cost: minimal
- Expected ROI: less tangible

Business Cases

- Kielikone/Rautaruukki
- PricewaterhouseCoopers
  - Different approaches for different niches
- Is there any way to compete with Systran?
  - DaimlerChrysler

TranSmart @ Rautaruukki

- Finnish-English Transfer (Kielikone)
  - Plus terminology management and source language checking
- Domain: Business communications for specialized steel industry (2.7B Euro)
- Coverage: core + domain
- Speed: 4000 words/minute
- Customization: 6-8 months
- Key requirement: terminology standardization

TranSmart: MT Output Markup

The worst general environmental effects of the sulphur dioxide are the acidification of systems. In large cities the irritation of the health of the human being by mainly the irritation of the skin and the lungs monitors the amount of the sulphur dioxide in the air of the factory area. The monitoring system consists of a same size of different measuring stations. Mainly the declining of the sulphur content of concentrates and the start-up of the production of the concentrates have affected a decrease in pollution in 1992. In addition to the steel mill for example, the steel mill of the air of Rautasu.

Translation not available
### TranSmart [2]

- **Success Factors**
  - User consortium (w/Nokia and Trantex) to involve users early
  - Iterative refinement to tune lexicon before full-scale launch
  - Despite early technical problem, deployment was a success

### TranSmart [3]

- **Benefits**
  - 500 technical and commercial users; covers 90% of translation domain
  - Intangible: consistent, correct use of terminology
  - Tangible: reduced cost in customer communication; “no need to use a business agent any more”

### PricewaterhouseCoopers

- **Global internal deployment of “imperfect but fast” MT for gisting**
- **Tool for business executives, not translators**
  - Desktop interface (e.g. web browser)
  - Auxiliary tools (dictionary, translation memory) not an important factor

### PWC [2]

- **Three potential vendors evaluated on pricing, accuracy and readability**
- **Systran was selected**
  - “State-of-the-art” translations
  - Willing to defer payment until system is up and running
  - (Other vendors wanted $Ms up front)

### PWC [3]

- **Customization: 3 months**
  - Incorporate existing glossaries
  - Implement authentication, access to MT inside existing infrastructure
  - Integration via ASP hosted @ Systran
  - Initially deployed in Spain, other countries added later

### PWC [4]

- **Pricing**
  - Desktop license: $1000 per seat
  - Corporate service: ranges between
    - $13,500 (100 users, 5 language pairs)
    - $77,200 (unlimited users, 5 pairs)
PWC [5]

• Benefits
  – Multilingual content search (large-scale CLIR not available commercially)
  – Fast gisting by managers to identify material to be translated by humans
  – Allows non-native specialists to join work teams
  – Access to translated terminology for document creation

PWC [6]

• Critical Success Factors / Lessons
  – Buy-in from professional translators to evaluate and customize
  – Business case: individual productivity enhancement vs. quantifiable savings
  – Setting realistic expectations
  – Ubiquitous desktop interface
  – Global licensing
  – Customization on a per-country basis

Why Systran?

• Why does the oldest technology have the most business success?
  – Continuous development for 30+ years
  – Extensive coverage (36+ pairs)
  – Very large grammars and dictionaries
  – Translation quality
  – Recent re-design of architecture for easier integration and modular re-use

Why Systran? [2]

• DaimlerChrysler selected Systran because of:
  – Ease of integration
  – Low performance costs
  – Language pairs
  – Translation quality

Two Business Models

• “Deeply customized solution for a niche market” (Kielikone)
  
• “One size fits all” (Systran)

If you don’t have a specialized niche market, it will be hard to compete with Systran… the Microsoft of the MT world?

Good Product vs. Good Project Management

• What are the “real” requirements?
• Customer involvement, “ownership”
• Project management
• Risk management
• Managing evolving requirements
• Technology transfer
• Goal: “No surprises”
Future

- Rapid deployment reduce time-to-market
- Improved quality reduce human cost
- Better support tools reduce support cost

KANT vs. KANTOO

- Improved maintenance
  - better modular design
  - more easily maintained language (C++ instead of Lisp)
- Better support tools
  - LMT (source terminology)
  - LTD (target terminology)
  - KMT (debugging & knowledge edit)

Return on Investment

- Development cost
- Development time
- Yearly cost of manual translation
- Yearly software/hardware support
- MT speedup factor
- ROI: How quickly can we recoup our investment from the savings we’ll achieve with MT?

ROI Examples

- 2 years to develop a $1M system
- Translator cost is $50/page
- Translation volume varies from 10K to 100K pages/year
- Support cost is $250,000/year
- Assume MT speedup factor is 2:1

Scenario I

<table>
<thead>
<tr>
<th>Year</th>
<th>Development Cost</th>
<th>Support Cost</th>
<th>Translation Volume</th>
<th>Translation Cost/Page</th>
<th>Speedup Factor</th>
<th>Adjusted Cost/Page</th>
<th>Savings vs Manual</th>
<th>Translation Cost</th>
<th>Total Cost</th>
<th>ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>500,000</td>
<td>0</td>
<td>100,000</td>
<td>50</td>
<td>1.00</td>
<td>50</td>
<td>-500,000</td>
<td>5,000,000</td>
<td>5,500,000</td>
<td>-500,000</td>
</tr>
<tr>
<td>Year 2</td>
<td>500,000</td>
<td>0</td>
<td>100,000</td>
<td>50</td>
<td>1.00</td>
<td>50</td>
<td>-500,000</td>
<td>5,000,000</td>
<td>5,500,000</td>
<td>-500,000</td>
</tr>
<tr>
<td>Year 3</td>
<td>0</td>
<td>0</td>
<td>100,000</td>
<td>50</td>
<td>0.50</td>
<td>25</td>
<td>0</td>
<td>2,250,000</td>
<td>2,250,000</td>
<td>2,250,000</td>
</tr>
<tr>
<td>Year 4</td>
<td>0</td>
<td>0</td>
<td>100,000</td>
<td>50</td>
<td>0.50</td>
<td>25</td>
<td>0</td>
<td>2,250,000</td>
<td>2,250,000</td>
<td>2,250,000</td>
</tr>
<tr>
<td>Year 5</td>
<td>0</td>
<td>0</td>
<td>100,000</td>
<td>50</td>
<td>0.50</td>
<td>25</td>
<td>0</td>
<td>2,250,000</td>
<td>2,250,000</td>
<td>2,250,000</td>
</tr>
</tbody>
</table>

Scenario II

<table>
<thead>
<tr>
<th>Year</th>
<th>Development Cost</th>
<th>Support Cost</th>
<th>Translation Volume</th>
<th>Translation Cost/Page</th>
<th>Speedup Factor</th>
<th>Adjusted Cost/Page</th>
<th>Savings vs Manual</th>
<th>Translation Cost</th>
<th>Total Cost</th>
<th>ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>500,000</td>
<td>0</td>
<td>100,000</td>
<td>50</td>
<td>1.00</td>
<td>50</td>
<td>-500,000</td>
<td>5,000,000</td>
<td>5,500,000</td>
<td>-500,000</td>
</tr>
<tr>
<td>Year 2</td>
<td>500,000</td>
<td>0</td>
<td>100,000</td>
<td>50</td>
<td>1.00</td>
<td>50</td>
<td>-500,000</td>
<td>5,000,000</td>
<td>5,500,000</td>
<td>-500,000</td>
</tr>
<tr>
<td>Year 3</td>
<td>0</td>
<td>0</td>
<td>100,000</td>
<td>50</td>
<td>0.50</td>
<td>25</td>
<td>0</td>
<td>2,250,000</td>
<td>2,250,000</td>
<td>2,250,000</td>
</tr>
<tr>
<td>Year 4</td>
<td>0</td>
<td>0</td>
<td>100,000</td>
<td>50</td>
<td>0.50</td>
<td>25</td>
<td>0</td>
<td>2,250,000</td>
<td>2,250,000</td>
<td>2,250,000</td>
</tr>
<tr>
<td>Year 5</td>
<td>0</td>
<td>0</td>
<td>100,000</td>
<td>50</td>
<td>0.50</td>
<td>25</td>
<td>0</td>
<td>2,250,000</td>
<td>2,250,000</td>
<td>2,250,000</td>
</tr>
</tbody>
</table>
### Scenario III

<table>
<thead>
<tr>
<th>Development Cost</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500,000</td>
<td>500,000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Support Cost</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>250,000</td>
<td>250,000</td>
<td>250,000</td>
<td>250,000</td>
<td>250,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Translation Volume</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Translator Cost/Page</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speedup Factor</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adjusted Cost/Page</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Savings vs Manual</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-500,000</td>
<td>-500,000</td>
<td>-500,000</td>
<td>-500,000</td>
<td>-500,000</td>
<td>-500,000</td>
<td>-500,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Translation Cost</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500,000</td>
<td>500,000</td>
<td>50,000</td>
<td>50,000</td>
<td>50,000</td>
<td>50,000</td>
<td>50,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total Cost</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,000,000</td>
<td>1,000,000</td>
<td>500,000</td>
<td>500,000</td>
<td>500,000</td>
<td>500,000</td>
<td>500,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ROI</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-500,000</td>
<td>-500,000</td>
<td>-400,000</td>
<td>-400,000</td>
<td>-400,000</td>
<td>-400,000</td>
<td>-400,000</td>
</tr>
</tbody>
</table>

### Questions?