Changing the Perspective on Machine Translation

Radu Soricut, Senior Research Scientist, SDL
What the MT community was concerned with in the past (2002-2012)

- SMT
- BLEU
  - Incremental Bleu improvements
- Technology: PBMT; SBMT; STG;...
  - Domain adaptation
  - Quality Estimation
  - Parallel Data Identification
  - SMT vs. RbMT vs. Hybrid
  - TM Repositories
  - Customization
    - Value-driven MT Apps and Use Cases

(SDL logo: Because Business is Global)
What matters to customers and end-users?!

• Customers and end-users don’t care if a system uses RbMT, SMT, or “hybrid” technology...
• They don’t care about BLEU...
• They don’t care about incremental advances...
• They don’t think in terms of data availability...

They care about value
Lots more value can be created when MT is part of an enterprise-based ecosystem.
Lots more value can be created when MT is part of an enterprise-based and a global ecosystem.
The Equation of Statistical Machine Translation

MT Quality

Amount of data

Usage/Feedback data

Vertical, domain-specific data

Better algorithms

More data

Better algorithms

Vertical, domain-specific data

More data

Usage/Feedback data
What you get with MT as part of an ecosystem
Massive TM Asset Exploitation

Billions of Web pages

Sentence-aligned parallel corpus

Usage

User feedback

TM assets

Eng>Jpn 2.74 3.08
Eng>Chi 1.86 1.94
Eng>Dan 2.72 3.26
Eng>Pol 1.9 2.39

Jan-11  
Jan-12
Adaptation to specific customer usage patterns

Enterprise needs/usage

Data Selection

Parallel Data

Training Data

Usage

TM assets

User feedback

Eng>Dan

3.25

3.58

Eng>Pol

2.39

3.6

Generic

Travel

SDL
Because Business is Global
Auto-awareness of MT failures

Usage - User feedback
TM assets

Eng-Kor

Human Likert

Automatic Likert

TrustScore
Continuous integration of user feedback

BeGlobal Term and Brand Lists:
Add terms or phrases to these “dictionaries” in order to improve automated translations instantly.

Feedback and Suggest a Different Translation Widgets

Rate a translation

Suggest a different translation
Challenges for the future of MT (2012-?)

• **Scalability**: scale from 100s of customers to 10 000s of customers (each with unique characteristics & requirements)

• **Adaptation**: enterprises are continuously changing terminology, styles, etc.

• **Feedback**: how to automatically learn from feedback inside an MT ecosystem
Changing the perspective on MT

Instead of talking about technology components, we need to talk about ecosystems of value. That is what benefits customers and end-users the most.