Area Report
Machine Translation

Hervé Blanchon

CLIPS-IMAG

herve.blanchon@imag.fr

Outline

- ! Some numbers
- ! Addressed topics
- ! Papers' content
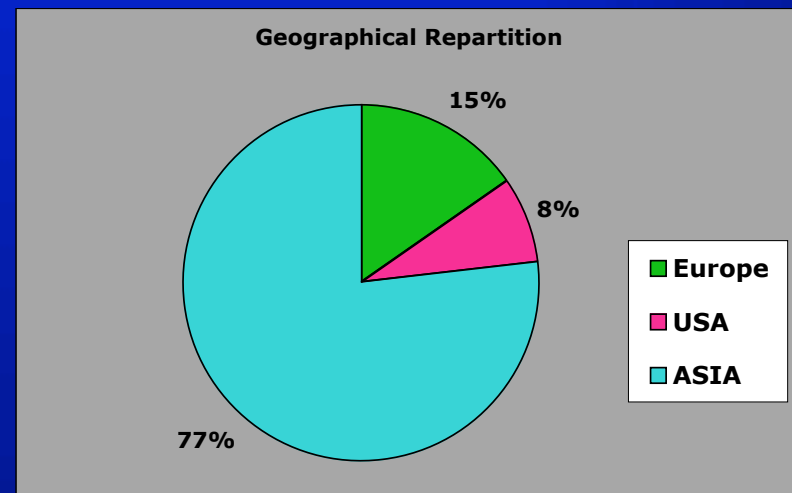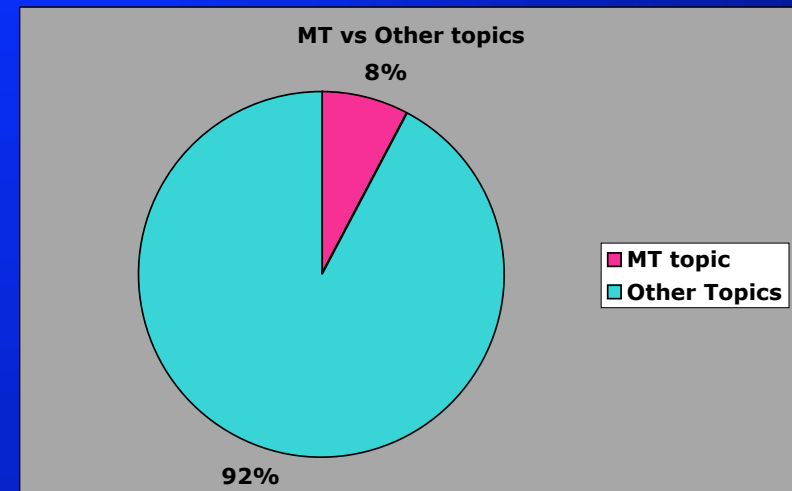- ! Questions
- ! My answers

Some Numbers

- ✦ **13 Papers** /170
  - – 12 paper in 4 sessions
  - – 1 paper in Learning
- ✦ Geographical distribution
  - – Europe: 2
    - • France: 1
    - • Germany+Spain: 1
  - – USA: 1
  - – Asia: 11
    - • China: 2
    - • Japan: 5
    - • Korea: 2
    - • Taiwan: 1

MT vs Other topics

8%

■ MT topic
■ Other Topics

92%

Geographical Repartition

15%

8%

■ Europe
■ USA
■ ASIA

77%

Some numbers

- Isn't this field under-represented?
  - 2 sub-committees for MT
    - MT, Asian Languages (11 papers)
    - MT, Western Languages (3 papers)
      - What is happening in Europe and the US?

- Is it due to the multiplication of conferences?

- Except ATR-SLT and I nothing on Speech-to-Speech Translation
  - Concurrence of ACL and ICSPL special sessions?

Addressed Topics

- Transfer-Based MT (with learning)
  - New language pair quick development
    - [Pinkham J. et Smets M.]
- Pivot-Based MT (STS Translation)
  - Shallow approach
    - [Blanchon H.]
- Example-Based MT
  - Evaluation
    - [Yao J. et al.]
  - Translation rules acquisition (learning)
    - [Echizen-ya H. et al.]

Addressed Topics

- **Statistical MT**
  - LM adaptation using MT to create new data (MT for Automatic Speech Recognition)
    - [Nakajima H. et al.]
  - Sentence alignment quality
    - [Kueng T.-L. et Su K.-Y.]
    - [Garcìa-Varea I. et al.]
  - Decoding (translation of a sequence of words)
    - [Watanabe T. et Sumita E.]
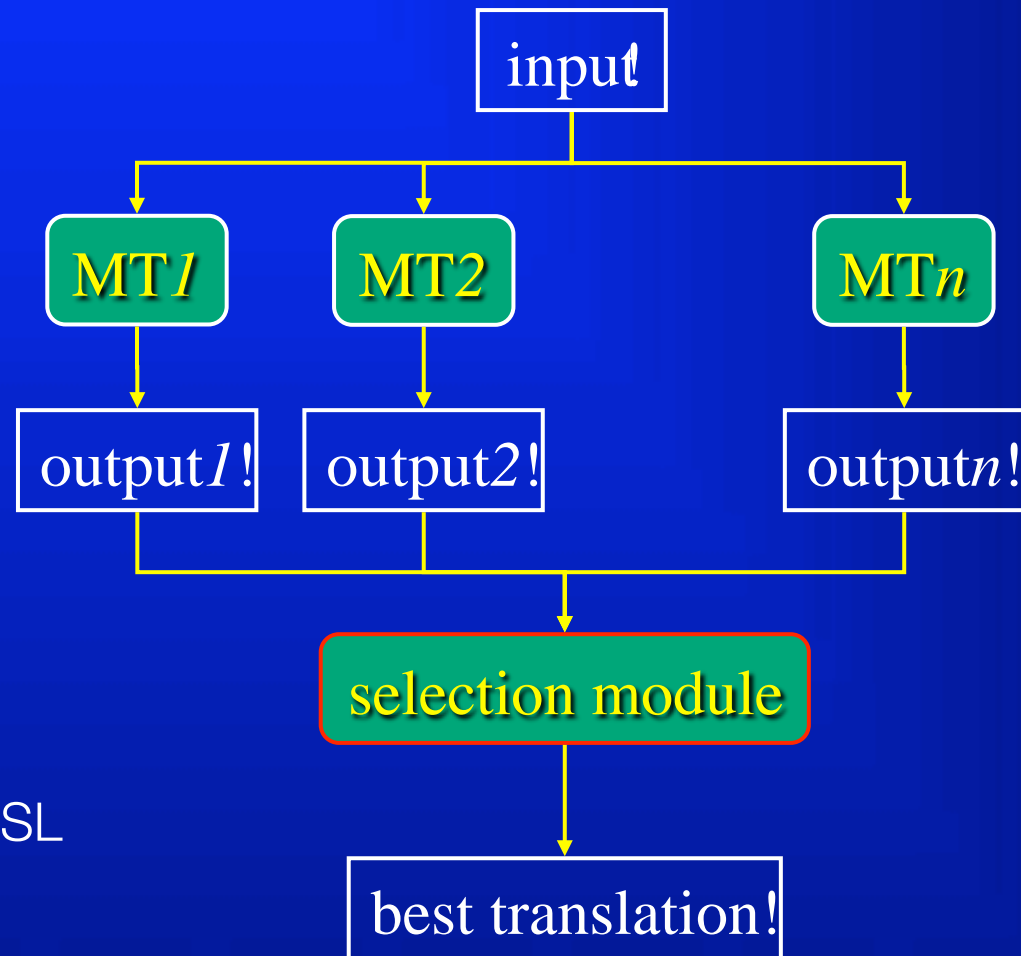
Addressed Topics

"New" Trends

✦ Multi-Engine MT

– Selecting the best translation among several

  • [Akiba Y. et al.]

– See also

  • VerbMobil project & ISL Lab. (CMU)
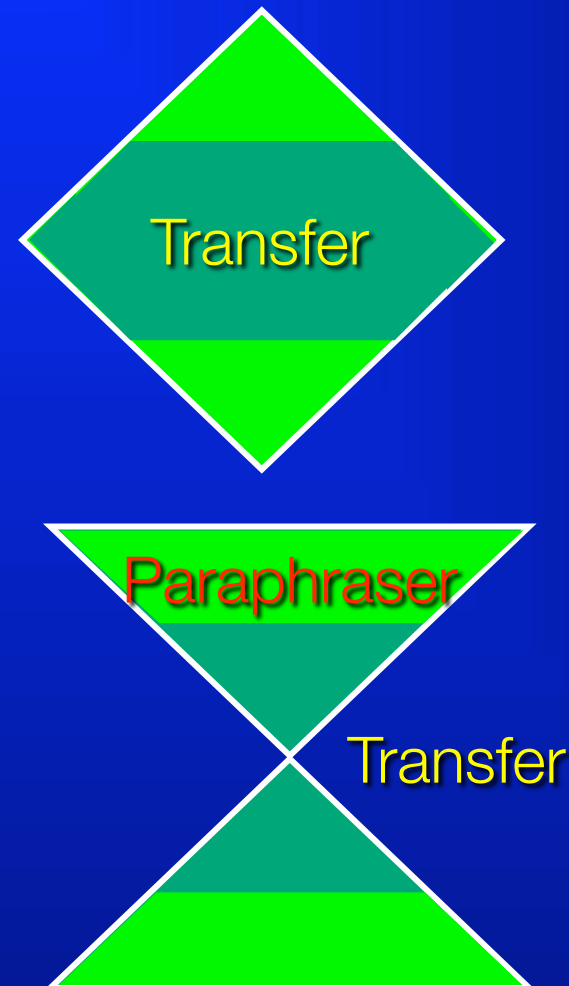
input

MT*1* MT2 MT*n*

output*1* output2 output*n*

selection module

best translation

Addressed Topics

✦! Sandglass MT
(paraphraser ↔ transfer)
  –! [Yamamoto K.]

Transfer

  –! If the transfer module fails it knows why
  –! Then it ask for a paraphrase giving constraints to the paraphaser

Paraphraser

Transfer

Addressed Topics

Old problem

✦ Translation selection (with statistics)
  – Getting better lexical transfer
    ● [Cao Y. & Li H.]
    ● [Kim Y.-S. & al.]
    ● [Lee H.A. & Kim G.C]

Papers' content

- ✦ Constants
  - – Measures
  - – Evaluations
- ✦ Wide range of results
  - – From fairly poor to high quality
- ✦ My impressions
  - – No real big steps
  - – Confirmation for the "statistics"
  - – No real vision (better results with more stats)

Questions

- Are the given results that good?
  - How would an ALPAC-like committee judge us?
  - How can we do a better job?
- Are the statistical methods the only way?
- New MT niches?
- Are we still aiming at FAHQT?
  - If so isn't there any possible intermediate steps for HQT?
- How keen are we in technological transfer?
  - The role of academic research?

Answers

- HQT through Interactive Disambiguation
  - Automatic disambiguation modules have to deliver a confidence measure
  - Added value of self-explained documents
    - Information for the semantic web
  - Moreover an Interactive Disambiguation module may also learn from the user!!!!!!
    - Word senses, Syntactic constructions

Answers

Do a better job

- Hybrid systems merging rule-based & statistical-based techniques
  - Statistics not at the level of strings but after some treatments and disambiguation
  - Using grammar for well handled phenomena
- Multi-engine systems
  - At the end of the process
  - At the beginning (self confidence level on an utterance)

Answers

✦!Adapt the MT technique to the targeted task/domain

– !Is there a better technique for a given task/domain?

- •!Narrow domain → ?
- •!Technical manuals → ?
- •!Translating the web → ?

– !We have to re-evaluate this question

Answers

- Translating the web?
  - Is translating the web the same task as translating a technical manual?
  - Have domain specific tuned systems?
- Speech-to-Speech Translation?
  - Ward, TMI-2002 MT Workshop
    - *"The current scenarios are science fiction and fairytales, the user interface is not taken into account"*
  - The story is about involving the user
- Aren't we trying to tackle more and more difficult problems without solving the previous ones?