Advances on Spoken Language Translation in the Quaero Program

K. Boudahmane, §B. Buschbeck, †E. Cho, ‡J. M. Crego,
*M. Freitag, †T. Lavergne, *H. Ney, †J. Niehues,
*S. Peitz, §J. Senellart, ‡A. Sokolov, †A. Waibel,
§T. Wandmacher, *J. Wuebker and ‡F. Yvon

IWSLT 2011, San Francisco
December 8, 2011

¶Direction générale de l’armement (DGA), France
†Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
‡LIMSI-CNRS, Orsay, France
*RWTH Aachen University, Aachen, Germany
§SYSTRAN Software, Inc.

http://www.quaero.org
2011 Spoken language translation (SLT) evaluation

- Annual project-internal evaluation conducted by DGA
- Goal:
  - evaluate market readiness and maturity of the developed technologies
- SLT eval builds upon previous year’s ASR eval
  - ASR Rover output as MT source
- Language pairs: German-French in both directions
- Data:
  - publicly available + internally collected
- Participants:
  - KIT, LIMSI, RWTH, SYSTRAN
Outline

► Evaluation framework
► Data description
► System descriptions
  ▶ KIT
  ▶ LIMSI
  ▶ RWTH
  ▶ SYSTRAN
► Results
► Summary
Evaluation framework

► Language pairs:
  ▶ German-French in both directions

► Conditions:
  ▶ (manual transcriptions)
  ▶ automatic transcriptions: ASR Rover, automatically segmented

► Evaluation data domain:
  ▶ mixture of broadcast news and broadcast conversation

► Scoring:
  ▶ two references produced by professional translators
  ▶ BLEU and TER
Data description

► Publicly available data
  ▶ bilingual and monolingual data from WMT 2010
    ACL 2010 Joint Fifth Workshop On Statistical Machine Translation

► Internally collected data (politics-news, UN documents)
  ▶ admin.ch
  ▶ project-syndicate.org
  ▶ bookshop.europa.eu
  ▶ presseurop.eu
  ▶ arte.tv
## Statistics for internally collected data

### Training data:

<table>
<thead>
<tr>
<th></th>
<th>German</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>16 K</td>
<td></td>
</tr>
<tr>
<td>Running words</td>
<td>5.3M</td>
<td>6.3M</td>
</tr>
<tr>
<td>Documents</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Running words</td>
<td>250 K</td>
<td>70 M</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documents</td>
<td>69 K</td>
<td></td>
</tr>
<tr>
<td>Running words</td>
<td>25 M</td>
<td></td>
</tr>
</tbody>
</table>

### Evaluation data:

<table>
<thead>
<tr>
<th></th>
<th>German-French</th>
<th>French-German</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Sentences</td>
<td>971</td>
<td>823</td>
</tr>
<tr>
<td>Running words</td>
<td>23K</td>
<td>21K</td>
</tr>
</tbody>
</table>
System description: KIT

- **Preprocessing**
  - **Training data:**
    - remove punctuation on source side
    - filter noisy data with SVM classifier
  - **Test data:**
    - smart casing to achieve higher coverage

- **In-house phrase-based decoder**
  - phrase extraction: Moses
  - 4-gram LMs with Kneser-Ney smoothing
  - parameters optimized for BLEU with MERT
System description: KIT

State-of-the-art extensions

- POS-based short-range reordering (Rottmann and Vogel, TMI 2007)
- POS-based long-range reordering (Niehues and Kolss, WMT 2009)
- phrase extraction from reordering lattice
- bilingual language model (Niehues et al., WMT 2011)
  - word-based
  - POS-based (German: RF tagger, French: LIA tagger)
System description: LIMSI

- **Preprocessing**
  - Test data:
    - remove partially recognized and repeated words

- **bilingual $n$-gram-based decoder **$N$-code**
    - monotone decoding
    - input: reordering lattice computed with FST using POS information
    - 4-gram LMs with Kneser-Ney smoothing
    - parameters optimized for BLEU with MERT
State-of-the-art extensions

- German POS-tagging with CRF-based tagger (Lavergne et al., ACL 2010)
- neural network language model (SOUL) (Le et al., ICASSP 2011)
  - 10-gram history size
  - applied in $n$-best list rescoring
System description: RWTH

▸ Preprocessing
  ▷ Training data:
    ◦ remove punctuation on source side
    ◦ add period at end of sentence

▸ In-house phrase-based decoder
  ▷ parameters optimized for BLEU with Downhill-simplex algorithm

▸ Hierarchical phrase-based decoder Jane (Vilar et al., WMT 2010)
  (http://www-i6.informatik.rwth-aachen.de/jane)
  ▷ parameters optimized for BLEU with MERT

▸ 4-gram LMs with Kneser-Ney smoothing
System description: RWTH

- State-of-the-art extensions
  - Triplet lexicon model (Hasan et al., EMNLP 2008)
  - Discriminative word lexicon model (Mauser et al., EMNLP 2009)
  - System combination (Leusch et al., WMT 2011)
System description: SYSTRAN

- Commercial MT system developed over decades
- Rule-based core engine with large-scale dictionaries
- Progressive integration of state-of-the-art MT techniques, e.g.
  - Statistical post-edition
  - Word sense disambiguation (WSD) models
  - Decision trees for POS disambiguation
- Combination of linguistic and statistical methods
System description: SYSTRAN

Rule-based translation is performed in 4 steps:

- Preprocessing
  - Segmentation
  - Normalization
  - Dictionary lookup

- Analysis
  - Morphological analysis
  - POS analysis
  - Named entity recognition
  - Syntactic dependency parsing

- Transfer
  - Application of transfer dictionaries and contextual disambiguation rules

- Synthesis
  - Syntactic rearrangement
  - Morphological generation
## 2011 Evaluation results

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU [%]</th>
<th>TER [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIT</td>
<td>18.4</td>
<td>70.4</td>
</tr>
<tr>
<td>LIMSI</td>
<td>13.4</td>
<td>71.0</td>
</tr>
<tr>
<td>RWTH</td>
<td>16.1</td>
<td>69.7</td>
</tr>
<tr>
<td>SYSTRAN</td>
<td>10.0</td>
<td>76.7</td>
</tr>
</tbody>
</table>

### German → French:

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU [%]</th>
<th>TER [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIT 2009</td>
<td>16.4</td>
<td>67.5</td>
</tr>
<tr>
<td>KIT 2010</td>
<td>17.7</td>
<td>66.1</td>
</tr>
<tr>
<td>KIT</td>
<td>18.9</td>
<td>68.0</td>
</tr>
<tr>
<td>LIMSI</td>
<td>17.0</td>
<td>68.7</td>
</tr>
<tr>
<td>RWTH 2009</td>
<td>12.0</td>
<td>70.1</td>
</tr>
<tr>
<td>RWTH 2010</td>
<td>17.3</td>
<td>66.7</td>
</tr>
<tr>
<td>RWTH</td>
<td>17.6</td>
<td>65.5</td>
</tr>
<tr>
<td>SYSTRAN</td>
<td>16.0</td>
<td>71.5</td>
</tr>
</tbody>
</table>

### French → German:

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU [%]</th>
<th>TER [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIT</td>
<td>18.4</td>
<td>70.4</td>
</tr>
<tr>
<td>LIMSI</td>
<td>13.4</td>
<td>71.0</td>
</tr>
<tr>
<td>RWTH</td>
<td>16.1</td>
<td>69.7</td>
</tr>
<tr>
<td>SYSTRAN</td>
<td>10.0</td>
<td>76.7</td>
</tr>
</tbody>
</table>
Rover from 2010 ASR eval as input for 2011 SLT eval

German-French in both directions

Public and internally collected data

KIT, LIMSI, RWTH

- statistical systems: phrase-based, hierarchical, $n$-gram-based
- state-of-the-art extensions developed in Quaero

SYSTRAN

- commercial rule-based engine

French $\rightarrow$ German: $+2.5\%$ BLEU since 2009
Thank you for your attention

Karim Boudahmane
Joern Wuebker

karim.boudahmane@dga.defense.gouv.fr
wuebker@cs.rwth-aachen.de

http://www.quaero.org