LIG English-French Spoken Language Translation System for IWSLT 2011
Benjamin Lecouteux, Laurent Besacier, Hervé Blanchon

To cite this version:
Benjamin Lecouteux, Laurent Besacier, Hervé Blanchon. LIG English-French Spoken Language Translation System for IWSLT 2011. IWSLT, 2011, San Francisco, United States. hal-02094743

HAL Id: hal-02094743
https://hal.archives-ouvertes.fr/hal-02094743
Submitted on 9 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract

This paper describes the system developed by the LIG laboratory for the 2011 IWSLT evaluation. We participated to the English-French MT and SLT tasks. The development of a reference translation system (MT task), as well as an ASR output translation system (SLT task) are presented. We focus this year on the SLT task and on the use of multiple 1-best ASR outputs to improve overall translation quality. The main experiment presented here compares the performance of a SLT system where multiple ASR 1-best are combined before translation (source combination), with a SLT system where multiple ASR 1-best are translated, the system combination being conducted afterwards on the target side (target combination). The experimental results show that the second approach (target combination) overpasses the first one, when the performance is measured with BLEU.

Introduction

We focus on the SLT task and on the use of multiple 1-best ASR outputs to improve translation. Two different approaches are proposed:

- source combination: multiple ASR 1-best are combined before translation
- target combination: multiple ASR 1-best are translated, before applying system combination on the target side

MT and SLT LIG systems in 2011

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU (x)</th>
<th>BLEU (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIG 2010 dev2010</td>
<td>0.2080</td>
<td>0.2179</td>
</tr>
<tr>
<td>LIG 2011 dev2010</td>
<td>0.2215</td>
<td>0.2311</td>
</tr>
<tr>
<td>Moses (T11)</td>
<td>0.2270</td>
<td>0.2367</td>
</tr>
<tr>
<td>Moses (T11)</td>
<td>0.2270</td>
<td>0.2367</td>
</tr>
<tr>
<td>Moses (T11)</td>
<td>0.2270</td>
<td>0.2367</td>
</tr>
</tbody>
</table>

SLT task:

- ASR output lowcased, tokenized and re-putated before translation
- True re-translation system for French: LM trained on uncased and uncased French data (Europarl+News+UN+Newsmonso: 24M sentences)
- Punctuation restored using hidden-ngram
- SMT-based recaser presented earlier

Source versus Target Combination

<table>
<thead>
<tr>
<th>System</th>
<th>WER%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sys 0 alone</td>
<td>0.1671/ 0.1602</td>
</tr>
<tr>
<td>Sys 1 alone</td>
<td>0.1609/ 0.1534</td>
</tr>
<tr>
<td>Sys 2 alone</td>
<td>0.1720/ 0.1664</td>
</tr>
<tr>
<td>Sys 4 alone</td>
<td>0.1770/ 0.1709</td>
</tr>
<tr>
<td>Target comb. (systems 42)</td>
<td>0.1770/ 0.1709</td>
</tr>
<tr>
<td>Source comb. (rover systems 420)</td>
<td>0.1851/ 0.1709</td>
</tr>
<tr>
<td>Target comb. (systems 420)</td>
<td>0.1851/ 0.1709</td>
</tr>
</tbody>
</table>

Source combination:

- classical ROVER weighted by the ASR WER
  \[ \alpha*\text{Sum(WordOcc)} + (1-\alpha)*\text{Sum(Confidence(W))} \]

Where \( \alpha=0.9 \) and confidence scores are empirically defined

Target combination:

- 500 best translated outputs generated from each ASR source system
- Moses option distinct
- N-best associated with a set of 13 features (10 TM, 1 distance-based, 1 LM, 1 word penalty)
  - Combined in several steps
  - Score combination weights optimized on a dev corpus (BLEU at the sentence level)
  - N-best resorted using SRLM nbest-optimize
- Once the optimized feature weights are computed independently for each ASR source:
  - N-best lists are turned into confusion networks (CN)
  - Features used to compute posteriors relatively to all the hypotheses in the N-best list
  - CN computed for each sentence and for each system
  - CN merged into a single one
  - ROVER is applied on the combined CN and generates a lowcased 1-best
- When 3 systems are available, the target combination is better than the source combination
- As more ASR systems (2, 3, 4) are added, the overall performance improves
- source+target combination show a slight BLEU degradation

Official results and Conclusion

- English-French MT updated on the new data without radical changes
- Several approaches to take advantage of multiple ASR system outputs
- Results show that combining translation hypotheses on the target language side lead to better results than combining ASR 1-best on the source side, before translation (0.4 BLEU improvement observed)