The LIGA (LIG/LIA) Machine Translation System for WMT 2011

Marion Potet, Raphaël Rubino, Benjamin Lecouteux, Stéphane Huet, Hervé Blanchon, Laurent Besacier, Fabrice Lefèvre

To cite this version:


HAL Id: hal-02088892
https://hal.archives-ouvertes.fr/hal-02088892
Submitted on 8 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.
The LIGA (LIG/LIA) Machine Translation System for WMT 2011

Marion Potet, Raphaël Rubino, Benjamin Lecouteux, Stéphane Huet, Hervé Blanchon, Laurent Besacier and Fabrice Lefèvre

Introduction

- Each group started by developing its own solution whilst sharing resources (corpora as provided by the organizers, aligned data, etc.) and acquired knowledge (current parameters, effect of the size of n-grams, etc.) with the other.

The final LIGA submission is a combination of two phrase-based translation systems with appropriate carefully-tuned setups.

LIA system

- Pre-processing:
  - truecased text
  - Reaccentuation of French words starting with a capital letter
  - 5-gram LM: mono-news-c + news-s
- Translation model: Training on 10M sentence pairs selected in news-c + eur + UN + giga

LIG system

- Pre-processing:
  - Lowercased text
  - Normalization of French euphonious ‘t’
- 4-gram LM: mono-news-c + news-s + mono-euro
- Translation model: Training on news-c + eur + UN

Results and Conclusion

Two statistical machine translation systems have been developed at different sites using MOSES and the final submission is the combination of these systems. Each individual system led to specific works.

The LIGA submission presented this year was ranked among the best MT system for the French-English direction.