FrameNet Annotations Alignment using Attention-based Machine Translation

Gabriel Marzinotto
1-Orange Labs, Lannion, France
gabriel.marzinotto@orange.com

Objectives
✓ Project FrameNet annotations into other languages using attention-based neural machine translation (NMT) models.
✓ Evaluate the approach using a multilingual BERT-based FrameNet parser

Translation and Label Projection Model
- NMT with encoder-decoder attention
- Use attention to align source and target languages
- Simple rules to project labels
  - Project labels only for verbs, nouns and adjectives
  - Complete Frame Elements (w/determinants and prepositions)
- Score label projections using the attention matrix and POS information
- Evaluate the projection : 19% labels (deletions) and 10% of duplicated labels

Human Translation vs Machine Translation
EN: We have a huge vested interest in it, partly because it’s education that’s meant to take us into this future that we can’t grasp.
PT-Human: Nos interessamos tanto por ela em parte porque é da educação o papel de nos conduzir a esse futuro misterioso.
PT-Machine: Temos um grande interesse, em parte porque é a educação que nos leva a esse futuro que não podemos compreender.

Semantic Parsing Models
- Semantic Parsing as a “Sequence Tagging Problem”
- 4 Layer bidirectional GRU on top of ML-BERT
- Parsing one trigger at a time
- Viterbi decoding and FrameNet Coherence filter to validate outputs

FrameNet Corpora et évaluation intrinsèque
✓ English FrameNet has much more frames we would like to exploit in French
✓ We translate SemEval-07-EN to obtain SemEval07-FR

Automatic Parsing Experiments
✓ Training parsers using translated corpora improves performances on the target language.
✓ Frame Identification benefits more than argument identification
✓ Ultimately combining the original corpus and its translation yields further improvement
✓ However, there are important differences between translated annotations and gold annotations

Conclusion
✓ Simple method to project FrameNet annotations into other languages using attention-based neural machine translation (NMT) models.
✓ A multilingual BERT-based FrameNet parser gives a strong baseline to evaluate how relevant the translated corpus is.
✓ Our method shows modest gains on English to French setting and shows that there is still room for improvement in the alignments.