Towards Recognizing Phrase Translation Processes

Experiments on English-French

Yuming Zhai, Pooyan Safari, Gabriel Illouz, Alexandre Allauzen, Anne Vilnat
LIMSI-CNRS, Univ. Paris-Sud, Univ. Paris-Saclay, Orsay, France

Introduction

1.EN: *and then finally, just* ***as they're running out of time***

↓ non-literal

2.FR: *et puis, à la fin, juste* ***quand le délai se rapproche***

↓ literal

3.EN: *and then finally, just* ***as the deadline approaches***

Literal translation vs. **other translation processes** (Vinay and Darbelnet, 1958; Newmark 1981, Chuquet and Paillard, 1989, Molina and Hurtado Albir, 2002, *etc.*)

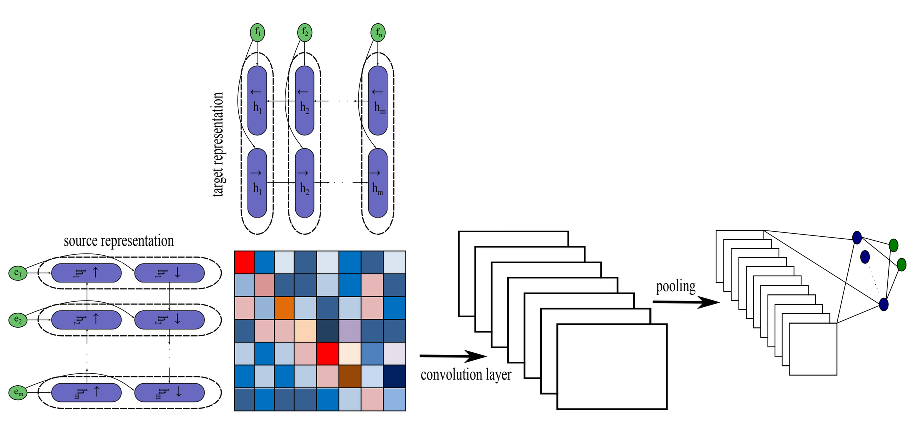
Translation Processes in Parallel Corpus

Translation Process and nb. instances	Typical example
Literal (3771)	*certain* ***kinds*** *of* → *certains* ***types*** *de*
Equivalence (289)	***back then*** → ***à l'époque*** ('at that time') ***Birds of a feather flock together.*** → ***Qui se ressemble s'assemble.*** ('(those) who resemble each other, assemble together.')
Transposition (289)	*unless* ***something changes*** → *à moins qu'****un changement ait lieu*** ('unless a change occurs')
Modulation (195)	***that scar has stayed with him*** → ***il a souffert de ce traumatisme*** ('he has suffered from this traumatism')
Mod+Trans (53)	*this is a completely* ***unsustainable*** *pattern* → *il est absolument* ***impossible de continuer sur*** *cette tendance* ('it is completely impossible to continue on this trend')
Generalization (86)	*as we* ***sit here*** *in ...* → *alors que nous* ***sommes*** *à ...* ('as we are at ...')
Particularization (215)	*the idea I want to* ***put out*** *is ...* → *l'idée que je veux* ***diffuser*** *c'est ...* ('the idea I want to spread is ...')

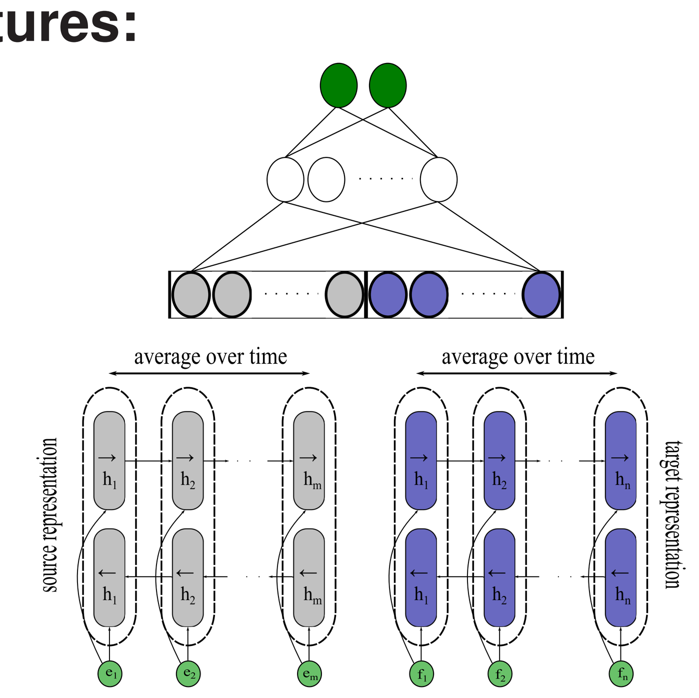
Corpus: TED Talks, manual annotation (Zhai et al., 2018)
Non_Literal class: 1127 instances
Transposition and *Mod+Trans* combined into *Contain_Transposition*

Automatic Classification: Two possible approaches

End-to-end Neural Network Architectures:



CNN: phrase alignment matrix classifier
input: character embeddings



MLP: averaged source and target representations classifier
input: character or word embeddings

Feature Engineering:

- Surface, PoS_tagging, Syntactic_analysis
- External_resource, Word_alignment

⟹ toolkit *Scikit-Learn*

Experimental Results

The evaluation is done by cross validation of five-folds.

End-to-end Neural Network Architectures using only embeddings:

Architecture	Accuracy	F1 (L)	F1 (NL)
Randomly initialized character embedding			
CNN	59.99%	0.60	0.60
MLP	71.16%	0.71	0.71
Pre-trained fasttext word embedding			
MLP	**71.25%**	0.71	0.71

Binary classification, balanced distribution

Architecture	Accuracy	Micro-F1	Macro-F1
Randomly initialized character embedding			
CNN	34.08%	0.34	0.20
MLP	40.74%	0.41	0.34
Pre-trained fasttext word embedding			
MLP	**43.22%**	0.43	0.34

5 non-literal classes

Statistical methods with feature engineering:

Distribution of classes	Classifier	Accuracy	Micro-F1	Macro-F1
Six classes				
six classes, with 3771 *Literal*	Dummy	60.76%	0.61	0.15
	RandomForest	**83.10%**	0.83	0.44
six classes, with 200 *Literal*	Dummy	18.92%	0.19	0.16
	RandomForest	57.04%	0.57	0.52
Two classes				
Literal (3) : *Non_literal* (1)	Dummy	65.84%	0.66	0.52
	RandomForest	**90.16%**	0.90	0.86
Literal (2) : *Non_literal* (1)	Dummy	56.43%	0.56	0.51
	RandomForest	88.85%	0.89	0.88
Literal (1) : *Non_literal* (1)	Dummy	53.19%	0.53	0.53
	RandomForest	**87.09%**	0.87	0.87
Five classes				
Five non-literal classes	Dummy	20.32%	0.20	0.18
	RandomForest	**55.10%**	0.55	0.47

Classification results under different configurations, using all features

Feature ablation study:

Binary classification: *word_alignment* contributes the most
Multi-class classification: *external_resource* contributes the least
PoS_tagging and *syntactic_analysis* contribute the most

Classifier (**RandomForest**)	average accuracy	average F1-scores	
binary classification (balanced distribution)	**87.09%**	0.87 (Literal)	0.88 (Non_literal)
five non-literal classes	**55.20%**	0.55 (micro average)	0.48 (macro average)

Results after feature ablation study

Configuration	average accuracy	average F1 (class1)	average F1 (class2)
L vs NL	**85.24%**	0.84	0.86
LE vs non-LE	75.32%	0.74	0.77
LET vs non-LET	**79.42%**	0.78	0.81

Results after grouping classes, every class has 549 instances

Conclusion and Perspectives

- First exploitation of translation processes in NLP
- New NLP task: automatic classification of translation processes
- Best experimental results: RandomForest with feature engineering
- Restrained by the limited dataset, but directions to follow are shown

+ Finer error analysis to help developing the annotation guidelines
+ Combine neural network architecture with feature engineering
+ Improve the classifier, extend the work to the pair English-Chinese
++ Better paraphrase extraction from bilingual parallel corpora (Bannard and Callison-Burch, 2005)

Contact Information

Code and dataset `https://github.com/YumingZHAI/ctp`
Personal Page `https://yumingzhai.github.io/`