Findings of the 2017 DiscoMT Shared Task on Cross-lingual Pronoun Prediction

UPPSALA UNIVERSITET

Sharid Loáiciga, Sara Stymne, Preslav Nakov, Christian Hardmeier, Jörg Tiedemann, Mauro Cettolo & Yannick Versley

Pronoun Translation

Pronoun translation is a problem for MT caused by:

- Mismatch in pronoun systems: differences in overtness, gender, number, case, formality, animacy, etc.
- Functional ambiguity: pronouns with the same surface form but different function.
- Translation diversity: a pronoun may not be translated as a pronoun in the target language.

Fill-in-the-gap Task

Given an input text and a translation with placeholders, replace the placeholders with pronouns.

SOURCE `If you ask for the happiness of the remembering self, it's a completely different thing.`

TARGET `Si vous réfléchissez sur le bonheur du "moi des souvenirs", REPLACE_11 est une toute autre histoire.`

CLASS `ce/c'`

Generalities

- The language pairs included are English→French, Spanish→English, and English↔German.
- Focus on subject pronouns, data is filtered accordingly.
- Baseline consists in 5-gram language models for each target language, trained on all training data and additional monolingual data from WMT.
- Macro-averaged recall is the official score.

Data

	Train				Dev & Test
	de-en	en-de	es-en	en-fr	all pairs
News Commentary v.9	X	X		X	
Europarl v.7	X	X	X	X	
TED talks	X	X	X	X	X

TED talks are particular with respect to pronoun use. Pronouns are frequent, including first and second person, but anaphoric references are not always clear.

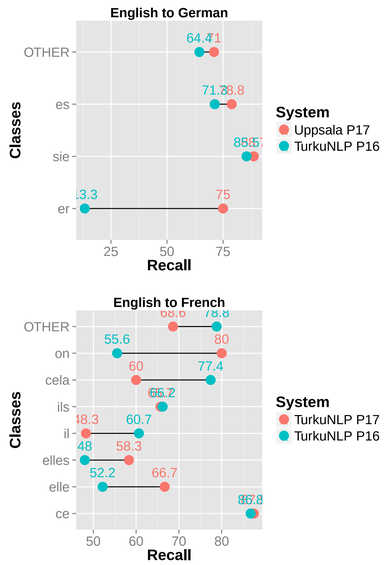
Submitted Systems and Results

Systems

5 teams of which 4 submitted primary (P) and contrastive (C) systems.

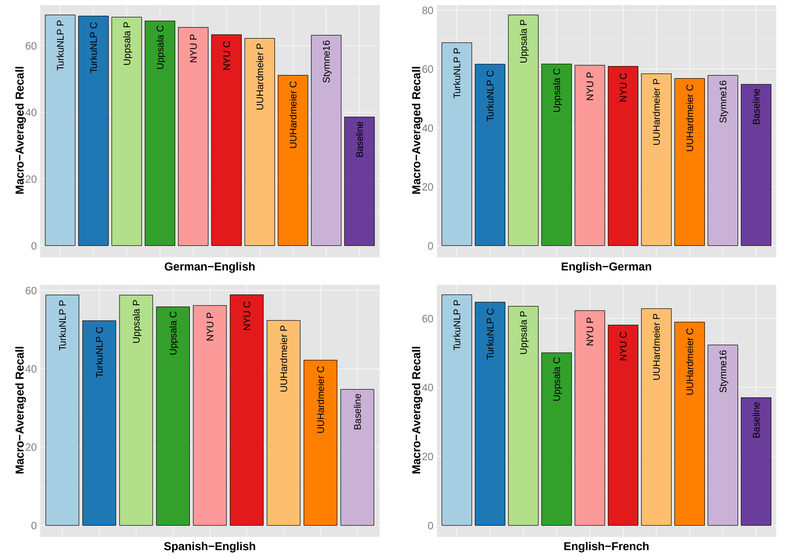
	TurkuNLP	NYU	Uppsala	UU-Hardmeier	UU-Stymne16
SVM					X
Neural networks	X	X	X	X	
-Convolutions	X			X	
-GRUs	X	X			
-BiLSTMs			X		
Source pronoun representation	X		X	X	X
Target POS tags	X		X		X
Head dependencies			X		X
Pre-trained word embeddings	X				
Source intra-sentential context	X	X	X	X	X
Source inter-sentential context		X		X	
Target intra-sentential context	X		X	X	X
Target inter-sentential context				X	

2016 and 2017: Best Systems Comparison



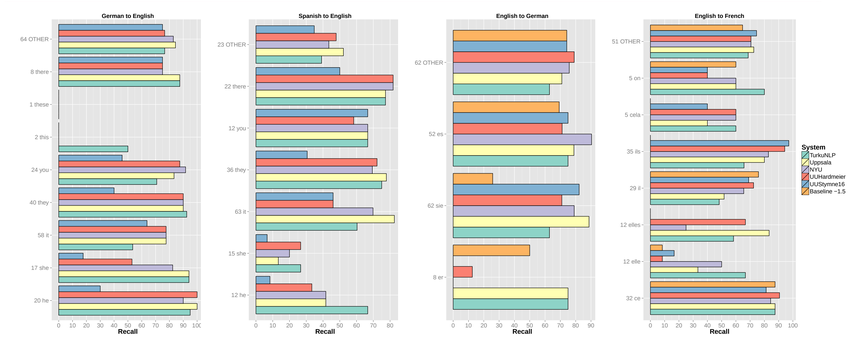
- Only English-German shows consistent improvements in all the classes.
- Improvement per class is very variable.

Results



- All systems outperformed the baseline. Spanish-English presents the lowest performance among the language pairs, while English-French obtained the best margin of improvement over the baseline.
- TurkuNLP, a bidirectional RNN with intra-sentential information, has the best macro-averaged recall overall.
- NYU submission is a full NMT system and their contrastive system ranked first for Spanish-English.
- Uppsala obtained the best macro-averaged recall and accuracy for English-German.
- UU-Hardmeier has the best accuracy for English-French.

Classes



Conclusions

- The shared task has made steady progress since its first edition in 2015. However, it is less clear that our understanding of pronoun translation has advanced.
- As in general MT, neural models have shown advantages for the task. n-gram language model baselines are no longer competitive.
- The task is far from solved, there is plenty of room for improvement.

This work was supported by the Swedish Research Council under project 2012-916 Discourse Oriented Statistical Machine Translation.