

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

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

8th September 2017



# Pronoun Translation

Machine translation problem caused by:

- ▶ Mismatch in pronoun systems: differences in gender, number, case, formality, animacy, etc.
- ▶ Null subjects: generating a pronoun in the target for which there is no pronoun in the source.
- ▶ Functional ambiguity: pronouns with the same surface form but different function.

English    Facit was a fantastic **company**. **They** were born deep in the Swedish forest, and **they** made the best mechanical calculators in the world.

German    Facit war ein großartiges **Unternehmen**. Entstanden tief im schwedischen Wald, bauten **sie** die besten mechanischen Rechenautomaten der Welt.

French    Facit était une **entreprise** fantastique, fondée dans la forêt suédoise. **Elle** fabriquait les meilleures calculatrices mécaniques au monde.

English     And among these organisms is a **bacterium** by the name of *Deinococcus radiodurans*. **It** is known to be able to withstand cold, dehydration, vacuum, acid, and, most notably, radiation.

German     Unter diesen Lebewesen existiert ein Bakterium namens *Deinococcus radiodurans*. Seine Resistenz gegen Kälte, Dehydratation, Vakuum, Säuren ist bekannt sowie insbesondere gegen Strahlung.

French     Et parmi ces organismes, il y a une **bactérie** appelée *Deinococcus radiodurans*. **Elle** est connue pour être capable de supporter le froid, la déshydratation, le vide, l'acide et, le plus notable, les radiations.

Spanish

[Arupa] también explicó que ella usó la retina de Thomas y su ARN para tratar de desactivar el gen que causaba la formación de tumores. Luego nos llevó al congelador y nos mostró las dos muestras que todavía conservaba. Dijo que las guardó porque no sabía cuándo podría conseguir más. Después de esto, [el personal de laboratorio] agasajó a Callum con [un regalo de cumpleaños]. Era el kit de laboratorio para niños. Y también le ofrecieron unas prácticas.

English

[Arupa] also explained that she is using Thomas's retina and his RNA to try to inactivate the gene that causes tumor formation. Then she took us to the freezer and she showed us the two samples that she still has. She said she saved it because she doesn't know when she might get more. After this, [the lab staff] presented Callum with [a birthday gift]. It was a child's lab kit. And they also offered him an internship.

We use this same word, depression, to describe how a kid feels when **it** rains on his birthday, and to describe how somebody feels the minute before they commit suicide.

A sense of belonging to the European Union will develop only gradually, as the EU achieves tangible results and explains more clearly what **it** is doing for people.

So in other words, I need to tell you everything I learned at medical school. But believe me, **it** isn't going to take very long.

**Pleonastic** We use this same word, depression, to describe how a kid feels when **it** rains on his birthday, and to describe how somebody feels the minute before they commit suicide. → **il**

**Nominal reference** A sense of belonging to the European Union will develop only gradually, as the **EU** achieves tangible results and explains more clearly what **it** is doing for people. → **elle, il**

**Event reference** So in other words, I need **to tell you everything I learned at medical school**. But believe me, **it** isn't going to take very long. → **cela, ça**

# Pronoun prediction

## What is the task?

SOURCE	If you ask for the happiness of the remembering self, <b>it's</b> a completely different thing.
TARGET	Si vous réfléchissez sur le bonheur du “moi des souvenirs”, <b>REPLACE_11</b> est une toute autre histoire.
CLASS	ce/c'

## Advantages:

- ▶ It defines set of possible translations (classes).
- ▶ It offers a controlled testing of different types of linguistic information.
- ▶ Explicit anaphora or coreference resolution is not necessary.

## More about the task

- ▶ 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.
  - ▶ Optimized with a penalty for shorter strings.
- ▶ 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.

# English → French example from the development dataset

---

ce OTHER

ce|PRON qui|PRON

It 's an idiotic debate . It has to stop .

**REPLACE\_0** être|VER un|DET débat|NOM idiot|ADJ

**REPLACE\_6** devoir|VER stopper|VER .|.

0-0 1-1 2-2 3-4 4-3 6-5 7-6 8-6 9-7 10-8

---

**pronoun class:** limited number of classes to predict

**original token:** not all original tokens are predicted

**source:** not modified

**lemmatized target:** discourages use of the target

**REPLACE\_X:** placeholder

**word alignments:** bidirectional word alignments

# Submitted Systems

- ▶ Turku NLP
- ▶ Uppsala
- ▶ NYU
- ▶ UU-Hardmeier
- ▶ Stymne16

Four new submissions for all language pairs, yielding also contrastive systems, and one comparative submission from a 2016 SVM system.

# Key characteristics of the submitted systems

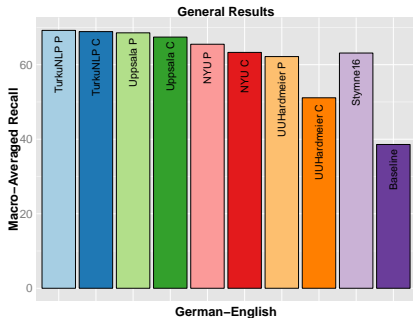
	TurkuNLP	NYU	Uppsala	UU-Hardmeier	UU-Stymne16
SVM					X
Neural networks	X	X	X	X	
-Convolutions	X			X	
-GRUs	X	X			
-BiLSTMs			X		
Source pronoun representation	X		X	X	X
Target POS tags	X		X		X
Head dependencies			X		X
Pre-trained word embeddings	X				
Source intra-sentential context	X	X	X	X	X
Source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

# Key characteristics of the submitted 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	

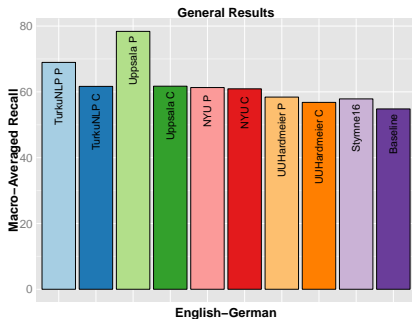
## Results

# German - English



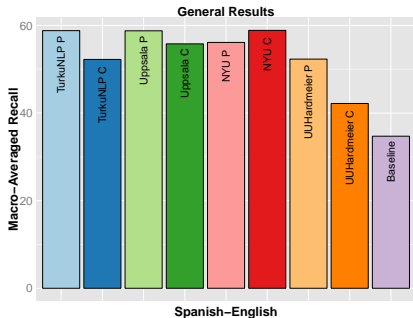
- ▶ Pair with the highest scores.
- ▶ TurkuNLP has the best macro-averaged recall (overall).
  - ▶ Bidirectional RNN with intrasentential information.

# English - German



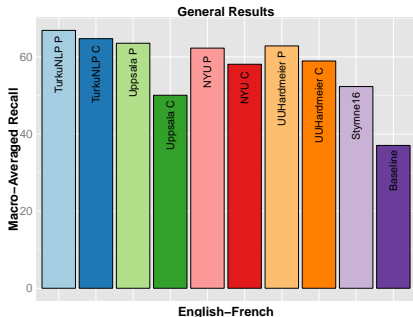
- ▶ Best ranking system is Uppsala Primary.
  - ▶ BiLSTM with head dependency information.
- ▶ There is a big difference between the first and second places.

# Spanish - English



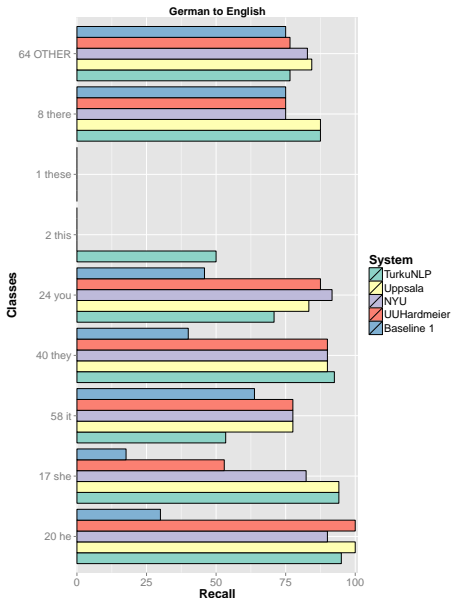
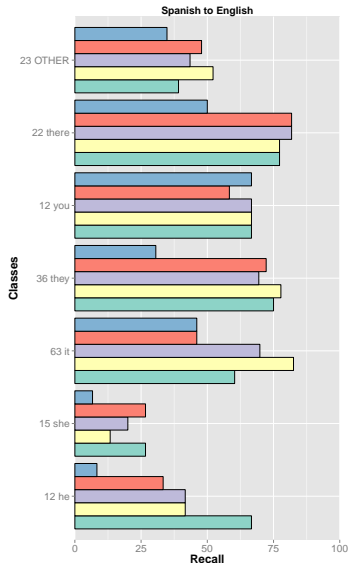
- ▶ Pair with the lowest scores overall.
- ▶ Best scoring systems are TurkuNLP Primary and NYU Contrastive.
  - ▶ Unlike the other teams, NYU submitted a full NMT system.

# English - French

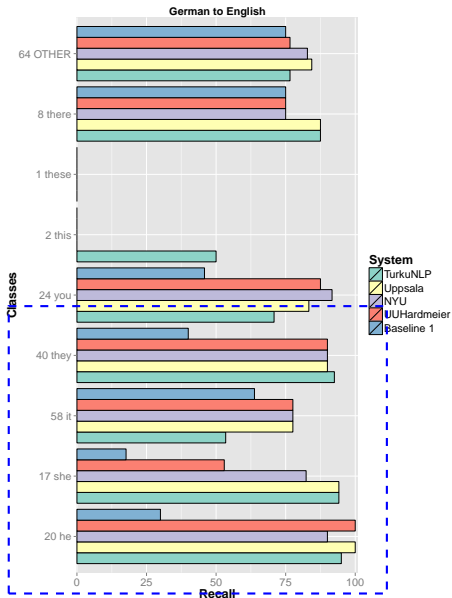
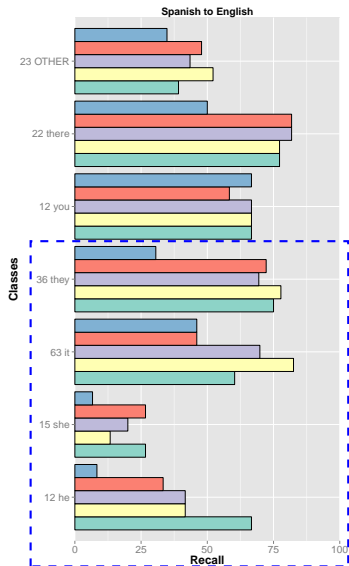


- ▶ Best margin of improvement over the baseline among the language pairs.
- ▶ TurkuNLP presents the best macro-averaged recall.

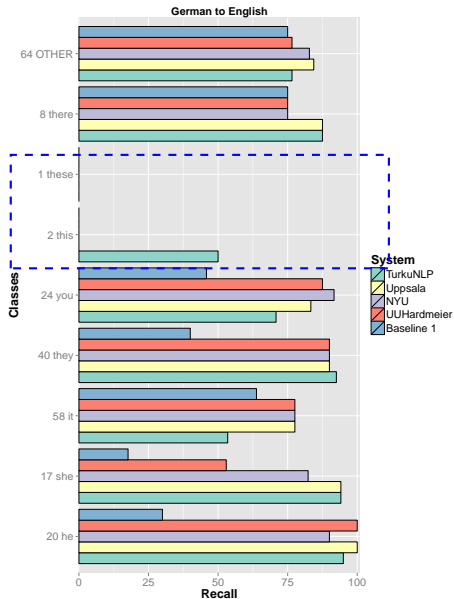
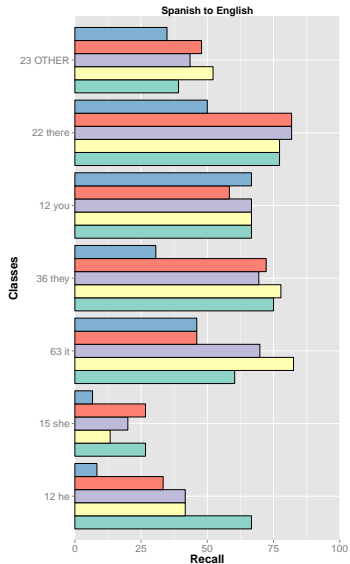
# What about the classes?



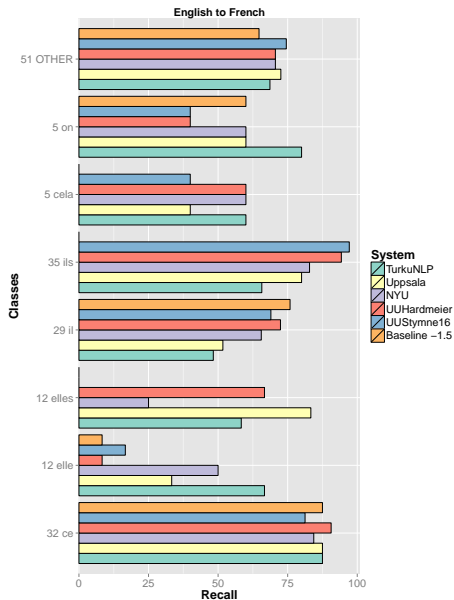
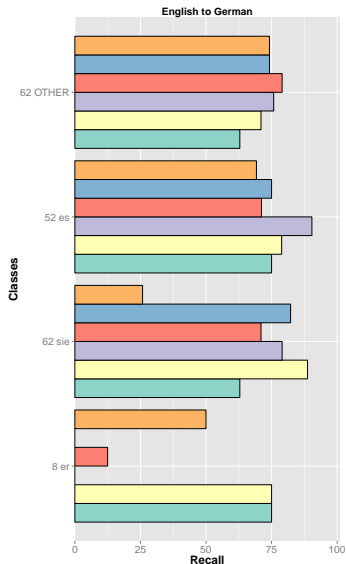
# What about the classes?



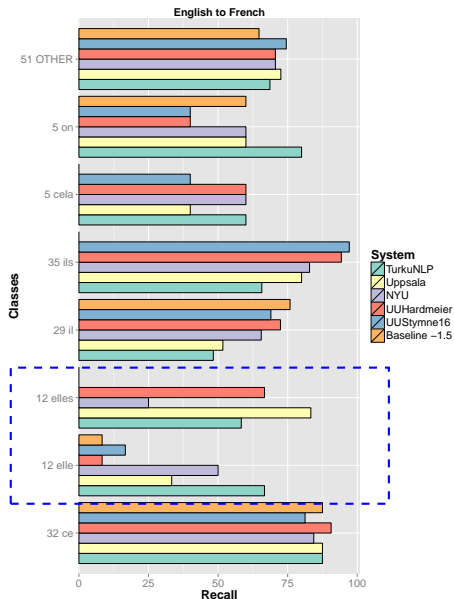
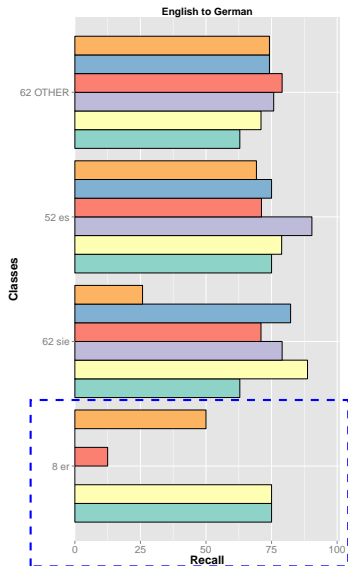
# What about the classes?



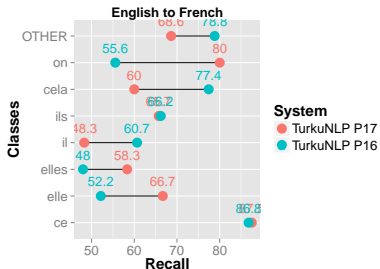
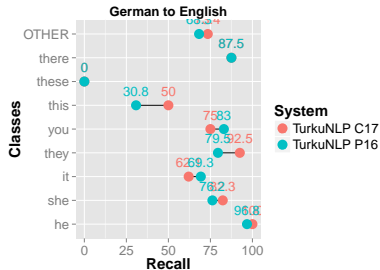
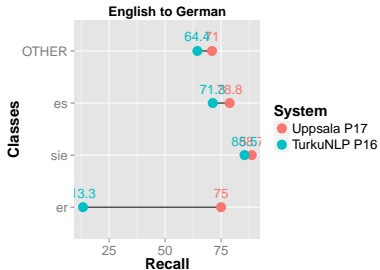
# What about the classes?



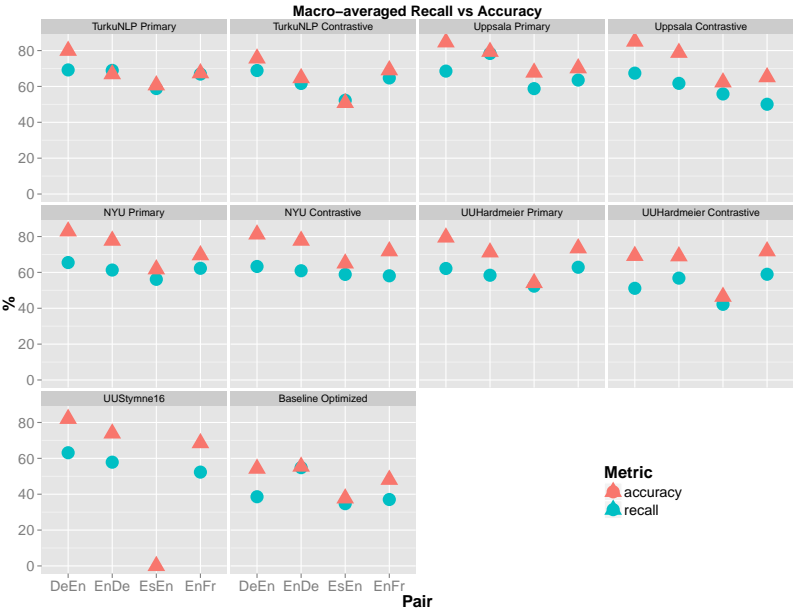
# What about the classes?



# Best 2017 and 2016 systems (Macro-avg. Recall)



# Metrics Comparison



## Questions yet to be answered

- ▶ Results have improved since 2015, but what have we learned about pronoun translation?
- ▶ Best performing systems rely on intra-sentential context, therefore, what is the importance of inter-sentential context?, and what is the best way to model it?
- ▶  $n$ -gram language models were meant to be competitive for SMT. Is this still a telling baseline?

# Conclusions

- ▶ The shared task winner this year is TurkuNLP.
- ▶ The shared task has made steady progress since its first edition in 2015. However, it is less clear that our understanding of the pronoun translation has advanced.
  - ▶ The NYU NMT system performed almost as well as specialized systems.
- ▶ As in general MT, neural models have shown advantages for the task.
- ▶ The task is not solved yet, there is plenty of room for improvement.

If you are interested in organizing this shared task next year,  
please let us know :)

Thank you!