



# CORBON 2017 SHARED TASK: Projection-based coreference resolution

Yulia Grishina

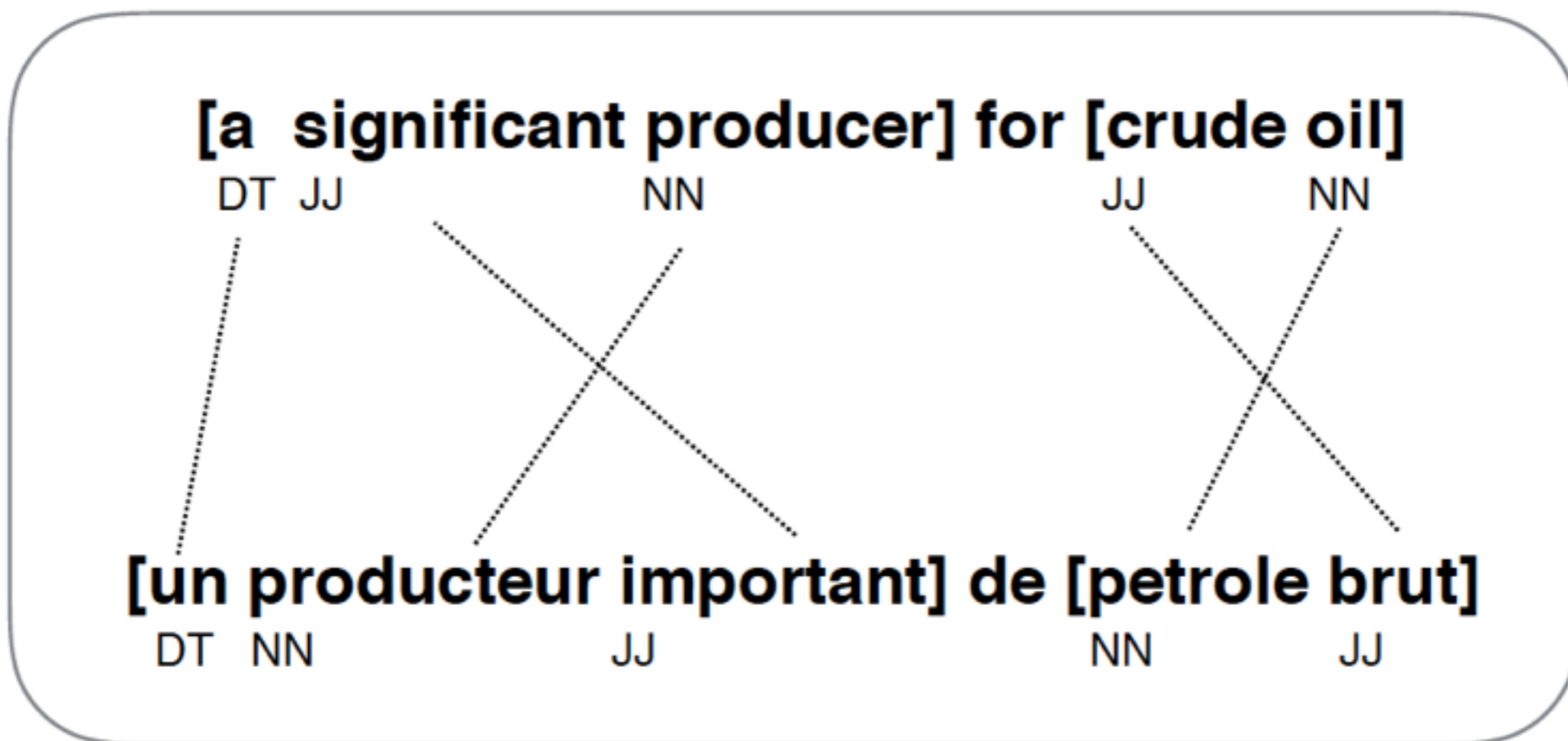
Applied Computational Linguistics  
FSP Cognitive Sciences  
University of Potsdam / Germany

# Task description

# Motivation

- rising interest in cross-lingual NLP
- annotation projection for coreference: (Postolache et al., 2006; Rahman and Ng, 2012; Grishina and Stede, 2015; Martins, 2015)
- multilingual evaluation tasks: SemEval 2010, CoNLL-2012
- lack of annotated multilingual resources

# Annotation projection



(Yarowski et al., 2001)

# Task setting

	Languages	Method	Resources
<b>Closed track</b>	German, Russian	annotation projection	parallel corpus; OntoNotes; automatic annotations
<b>Open track</b>		any	any

# Parallel corpus

- OPUS: NewsCommentary11
- already sentence-aligned
- tokenised using EuroParl tools
- labelled by Berkeley Entity Resolution System  
(Durrett & Klein, 2014)

# Test set

- manually annotated news (7) and short stories (3)
- OntoNotes compatible, except for:
  - only NPs annotated (no verbs)
  - appositions are part of markables (not a separate relation)
  - pronominal adverbs in German are annotated

# Task datasets

	Training set			Test set		
	#docs	#sents	#tokens	#docs	#sents	#tokens
<b>EN</b>			5,341,828	-	-	-
	5,749	221,844				
<b>DE</b>			5,404,568	10	413	8,753
<b>EN</b>			4,503,260	-	-	-
	4,869	188,761				
<b>RU</b>			4,290,892	10	413	8,092

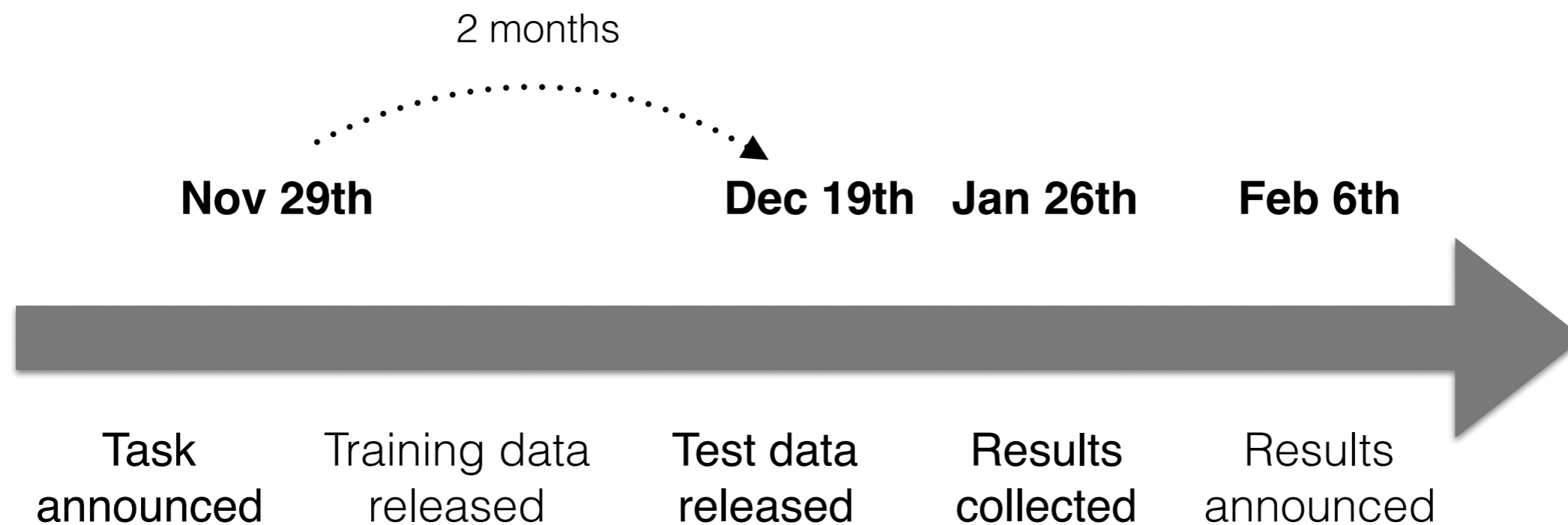


# Evaluation

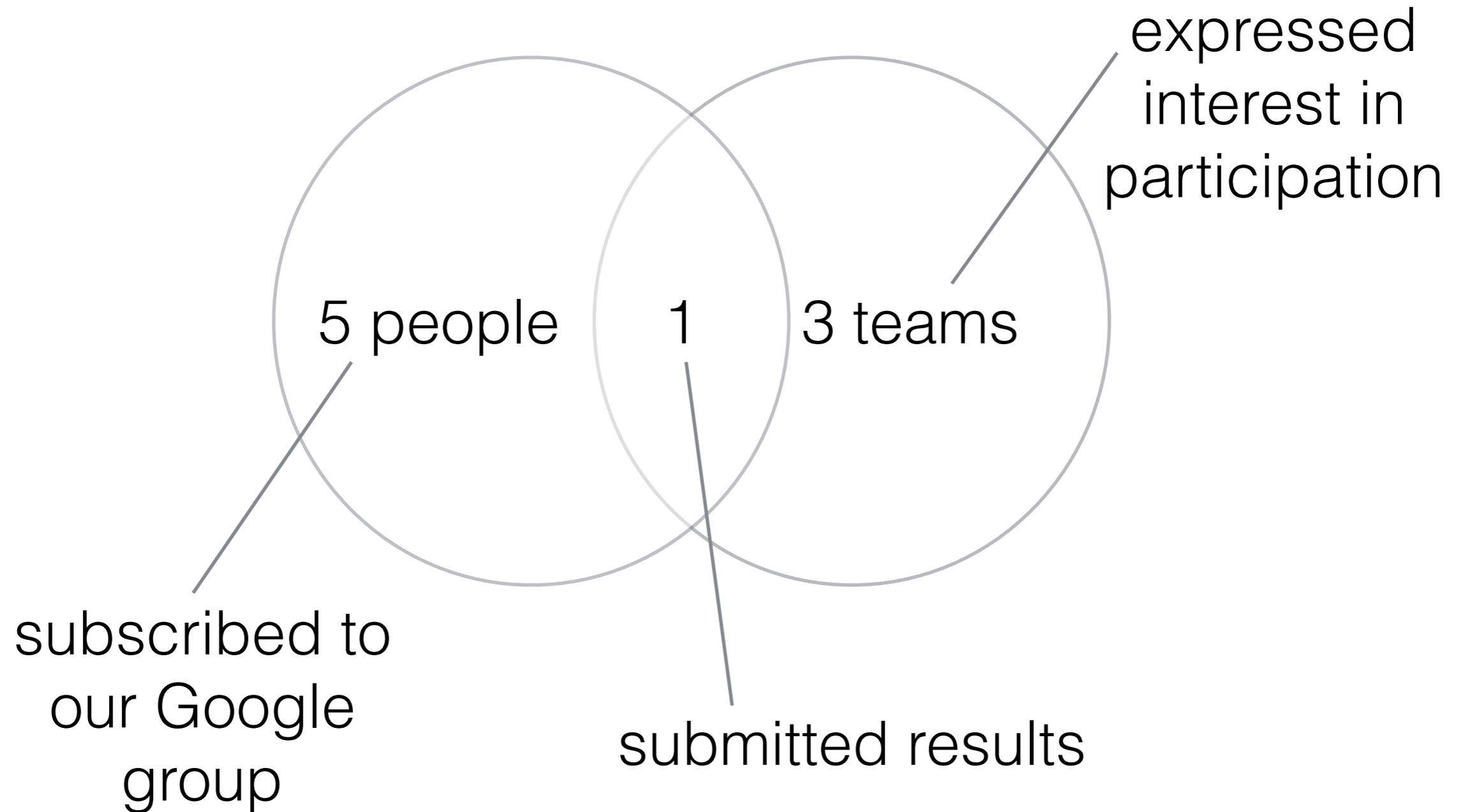
- CoNLL score: unweighted average of MUC, B-cubed and  $CEAF_e$
- CoNLL scorer: <https://github.com/conll/reference-coreference-scorers>
- No singletons

# Task progress

# Shared task timeline



# Task participants



# Results

	German			Russian			Official score
	P	R	F1	P	R	F1	
<b>Charles University, Prague</b>	62.5	20.3	29.40	59.56	21.2	30.94	30.17

# Results

- Precision: already able to achieve good P scores
- Recall: a bilingual dictionary or automatically induced paraphrases in order to retrieve missing coreference mentions on the target side

# Related work

- (Martins, 2015): F1=38.82 for Spanish and F1=37.23 for Portuguese
- (Grishina & Stede, 2015): P=68.0/82.1 and R=45.8/62.6 for German/Russian

# Outcomes

- first evaluation task on projection-based coreference resolution
- did not provide any manually annotated gold data, relied solely upon the automatic annotations
- the winning system: already able to achieve good Precision, but lower Recall
- data available at:

<https://github.com/yuliagrishina/CORBON-2017-Shared-Task>



# What's next?

- continue the tradition of evaluation tasks for coreference
- encourage more participants to take part
- promote multilinguality and cross-lingual methods
- Suggestions?

**thank you!**