

BRIDGING ANAPHORA RESOLUTION FOR THE RUSSIAN LANGUAGE

Anna Roitberg^{1,2}

Denis Khachko¹

¹IMPB RAS - Branch of KIAM RAS

²HSE RSU

ANAPHORA

```
graph TD; A[ANAPHORA] --> B[Direct anaphora  
Anaphor and Antecedent  
are coreferential]; A --> C[Indirect anaphora  
Anaphor and antecedent are  
not coreferential, but linked  
with some other relations];
```

Direct anaphora

Anaphor and Antecedent
are coreferential

Indirect anaphora

Anaphor and antecedent are
not coreferential, but linked
with some other relations

Direct Anaphora

A man came. **He** was tall.

A stone lay on the grass, the huge grey **rock**.

Very interesting texts! A have read the **collection**.

Indirect Anaphora

I came to a room. **The walls** were white.

There was a crime, **murder** still have not found.

Team played excited, the **couch** was proud of them.

Indirect anaphora = Bridging

An anaphoric relations between
two non-coreferential entities

Bridging anaphora

Usually:

- Meronymic relations
- Definite NPs (in article languages)

Sometimes:

- Non-meronymic semantic relations
- Without semantic restrictions
- All NPs
- Verb–argument relations

Genitive Bridging

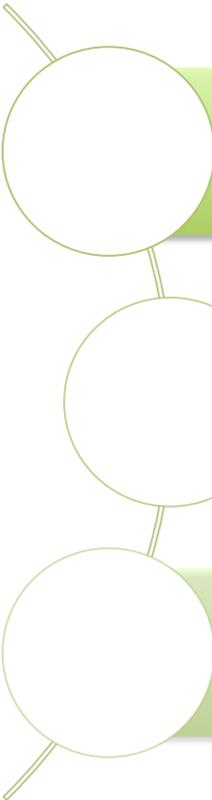
- A genitive dependent can be missing if its coreferential item appears in the previous text.

-> Genitive bridging

Там стоял грузовик с наклейками на **кабине**.

‘There was a huge truck with stickers on the **cab**’

What do we annotate?



Genitive bridging

In short texts (up to 10 sentences)

Bridging Anchors & Elements are NPs

*Противоестественная
разметка*

The RuGenBridge Corpus

Texts	339
--------------	------------

Tokens	69900
--------	-------

Bridging links	609
----------------	-----

Annotation

Automatic annotation:

- Lemmatization (FreeLing)
- Part of speech (FreeLing)
- Syntactic relations (MaltParcer)
- Semantic similarity (Word2Vec) – extended corpus ($5 * 10^6$ tokens)

Manual annotation

- Bridging relations (using BRAT tool)

RuGenBridge Corpus

TF/IDF

Part 1 (55 K tokens)

Part 2 (14 K tokens)

- Random Forest algorithm
- SciKit learn

RuGenBridge Corpus

TF/IDF

Part 1
(55 K tokens)

Classifier 1:
Bridging anchors
Bridging elements

Part 2 (14 K tokens)

RuGenBridge Corpus

TF/IDF

Part 1
(55 K tokens)

Classifier 1:
Bridging anchors
Bridging elements

Part 2
(14 K tokens)

Classifier 2:
Bridging pairs

Classifier 1: bridging elements & anchors.

Features

Semantic similarity to anchor\briding element

TF-IDF of word

Dystance in words from the beginning of the text

Lemma

Type of syntactic link from the word's head to word

Case

Syntax distance from word to root of senrence

Animacy

Bridging Anchors and Elements

Precision Maximized

	Bridging Element	Anchor
Precision	1	0.98
Recall	0.21	0.20
F-measure	0.35	0.61

PART 2
14 K tokens

**Bridging
Anchors**

**Bridging
Elements**

Formal Filter

6.5 K tokens

8.2 K tokens

Classifier 1

500 tokens

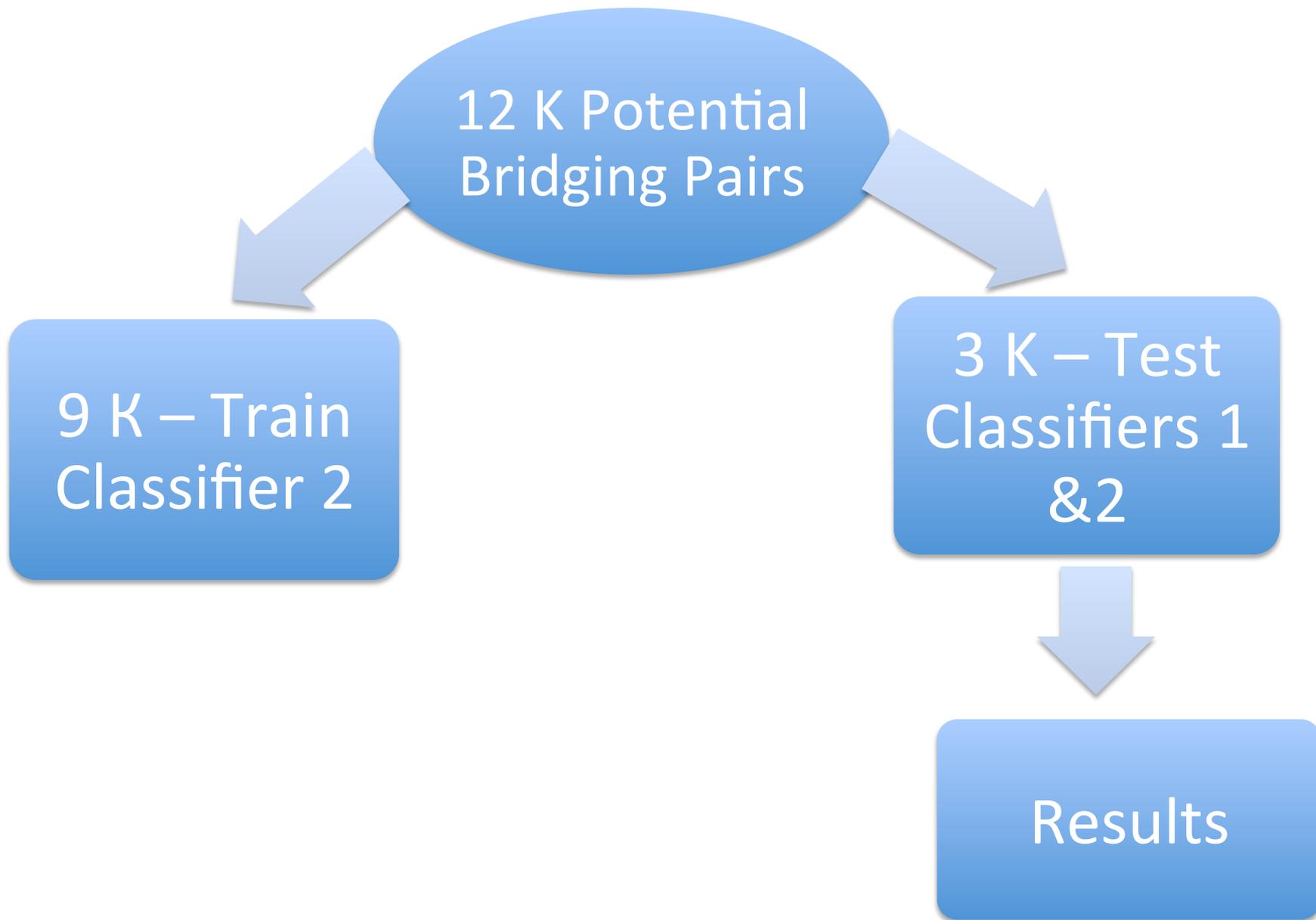
1.3 K tokens

500 Potential
Anchors

1.3 K Potential
Bridging Elements

Formal Filter

12 K
Potential
Bridging
Pairs



Bridging pairs features

Distance from bridging anaphor to anchor in words

Probability of anchor

Probability of bridging element

Lemma of bridging element

Syntactic distance from bridging anaphor to anchor

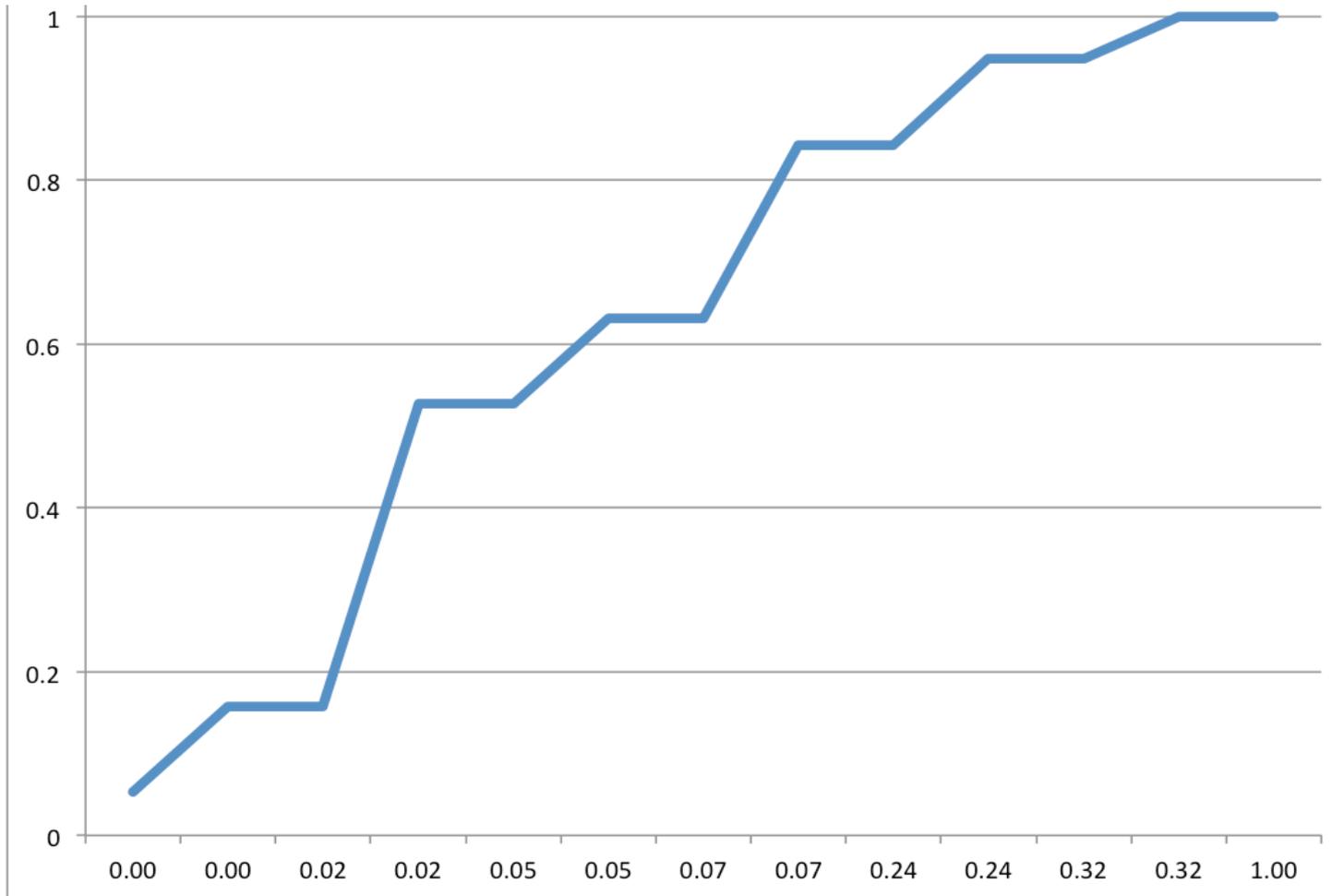
Lemma of anchor

Case of bridging anaphor

Case of anchor

Results. ROC for bridging resolution

AUC = 0.92



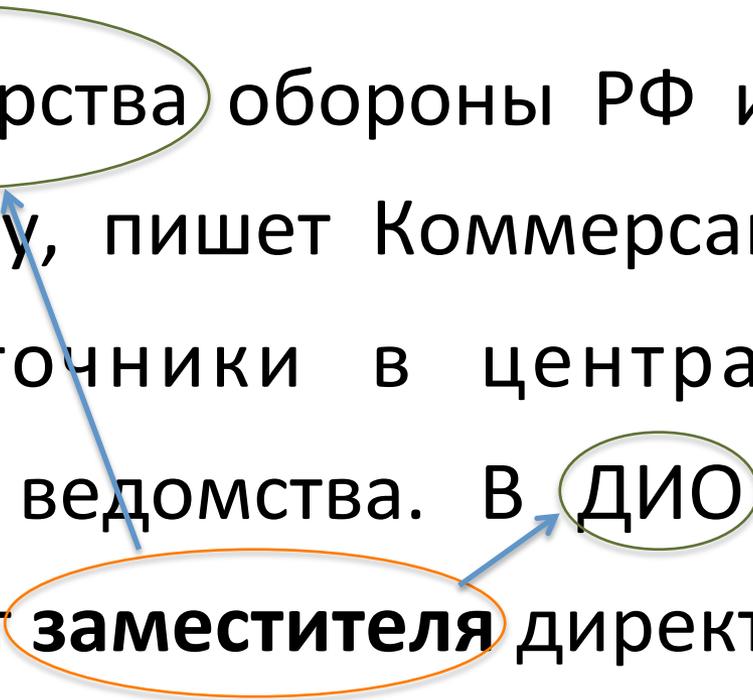
Bridging Resolution

F-Measure Maximized

Precision	0.58
Recall	0.73
F-measure	0.65

Examples

...**министерства** обороны РФ и был представлен коллективу, пишет Коммерсантъ со ссылкой на свои источники в центральном аппарате военного ведомства. В **ДИО** Rogozin-младший занял пост **заместителя** директора...



Examples

Первостепенное значение будет иметь улучшение потребительской и деловой уверенности за счет увеличения производства и стратегии снижения риска за пределами сырьевого рынка путем увеличения активов, услуг и клиентов. [Среднесрочные (N) перспективы развития] России ВБ оценивает как весьма низкие, но долгосрочные приобретают новые возможности.

Examples



...возмущен позицией главы СБУ
передавать официальные письма
через своего **пресс-секретаря** ...

A blue arrow points from the word 'пресс-секретаря' (circled in orange) to the word 'главы' (circled in green).

Examples



... которого подозревают в убийстве сбитого над сирийской приграничной зоной российского бомбардировщика Су-24, заявил, что не отдавал приказа об убийстве пилота. По словам Челика, он приказал своим подчиненным взять российского летчика в плен. Адвокат подозреваемого отметил, что его клиент берет на себя ответственность за гибель россиянина исключительно как командир своей группы. У Челика нет никакой ненависти ни к погибшему пилоту, ни к российскому народу, подчеркнул представитель подозреваемого в убийстве.

Examples



О реформе президент **России** Владимир Путин объявил на встрече с главой **МВД** Владимиром Колокольцевым, главой ФСКН Виктором Ивановым, командующим Внутренними войсками **МВД** Виктором Золотовым и заместителем главы **ФМС** Екатериной Егоровой.

Examples



в Instagram Папы Римского появилась
фотография понтифика,
обнимающего двух девочек,
предположительно, с синдромом
Дауна, с желто-голубой лентой в
руках.