

# Bridging Corpus for Russian in comparison with Czech

ANNA ROITBERG<sup>1</sup>, ANNA NEDOLUZHKO<sup>2</sup>

<sup>1</sup>IMPB of IPM RAS School of Linguistics HSE RSU, Russian Federation

<sup>2</sup>Charles University in Prague, Czech Republic

**Direct anaphora** - anaphor and antecedent are referring to the same discourse entity.

*There was an **apple** at the table, I gave it to John.*

**Bridging anaphora** - anaphor (*bridging element*) is associated to antecedent (*anchor*)

*There were some **fruits** on the table. John took an apple*

## Semantic Approach to Bridging Annotation - PDT

Bridging relations are annotated in the following cases:

1. PART-WHOLE & WHOLE-PART - part and a whole: *face – eyes*
2. SET-SUB & SUB-SET - set and its subsets or elements of the set: *a group of students – some students*  
FUNCT-P & P-FUNCT - an entity and a singular function on this entity: *company – director*
3. CONTRAST - coherence-relevant discourse opposites: *black flags – white flags*
4. REST – 6 subgroups (family members, event – argument, locality – inhabitant, etc.)
5. ANAPH - non-coreferential explicit anaphoric relation: *first world war – at that time*

## Syntactic Approach to Bridging Annotation - RuGenBridge

Bridging relations are annotated if two elements (an anchor/antecedent and a bridging element/anaphor) can form a genitive construction, where the anchor is marked with the genitive case in Russian.

*V avtobuse nachalsya pozhar. **Voditel'** [avtobusa.Gen] sam potušil ogon'.*

*'The fire broke out in the bus. The **driver** [of the bus] put out the fire by himself.'*

*The driver* is the driver of *the bus* from previous sentence → *Driver* and *bus* are anaphorically related.

'Driver bus\_Gen' – is grammatical → *driver* → BRIDGE → *the bus*

We use the NON-GEN label if the construction 'bridging element + anchor\_Gen' is ungrammatical due to lexical constraints

*Pravitel'stvo **Rossijskoj Federacii** vneslo na rassmotrenie (...) Etot proekt takz'e sposobstvuet razvitiju Sibiri.*

*[The Government of the **Russian Federation** brings a bill (...) This bill also promotes the devel opment of Syberia.]*

*Syberia* → NON-GEN → *Russian Federation*      \**Syberia Russiam Federatiom\_Gen* - ungrammatical

Useful for automatic extraction

## The experiments - application of the RuGenBridge and PDT schemes

**Hypothesis** Pairs annotated in RuGenBridge form semantic classes similar to PDT

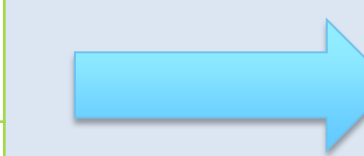
**Experiment 1.** Eight Russian texts annotated with PDT and RuGenBridge schemes independently:

Results: 69 PDT marks

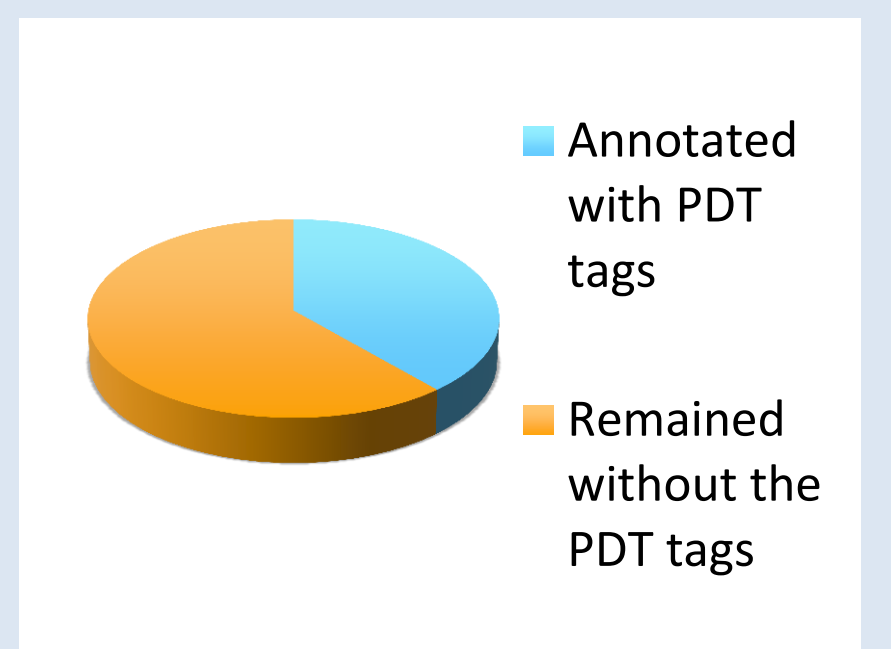
22 RuGenBridge marks **just one the same** → we catch different bridging

**Experiment 2.** We checked out all 430 cases of genitive bridging in the RuGenBridge corpus, and added PDT bridging marks to all pairs where it was possible.

Gen_Bridge annotation scheme	
BRIDGE	362
NON - GEN	8
<b>Total GenBridge</b>	<b>370</b>



PDT annotation scheme	
PART - WHOLE	73
FUNCT-P	55
SET - SUBSET	8
REST	16
<b>Total PDT</b>	<b>152</b>



**218 pairs remained without PDT tags:**

- 56 pairs {
  - Anchor - geographical name + bridging element - "something is located there" (*Moscow – hospitals*)
  - Anchor - geographical name + bridging element - "something is concerned with this geographic object" (*Russia – budget*)
- 162 pairs {
  - Object – its possessor (*flat – landlord*),
  - Object – something belonging to this object, but not the part of the object (*aerodrome – airplane*),
  - Expressions with the names of measures (*oil – barrel*),
  - Collocations, mostly deverbative nouns (*rates – increase*)

Anaphoricity of many pairs is given by situational relationships within the texts.

These relations do not have semantic nature, but they are not purely pragmatic.

**Conclusions:**

- Genitive bridging provides an opportunity to find out new functional types of bridging relations with respect to textual structure;
- Genitive bridging is more consistent than semantic annotation of bridging, because it is based on formal criteria and it does not require fixing a borderline between semantics and the world knowledge.

**Future work:**

- Annotate the existing corpus with two annotators and a supervisor, measure IAA;
- Extend the corpus and analyze bridging cases attested;
- Develop a system for genitive bridging resolution based on the information in the corpus